

明 細 書

生体関連事象間の相関データの可視化方法、解析法及びデータベース

技術分野

本発明は、生体関連事象間の相関データ、特に蛋白質、低分子化合物、DNA等の生体内物質間の相互作用情報や遺伝子の発現プロファイル等の視覚化方法に関する。また、本発明は、上記方法を取り入れたグラフィカルユーザーインターフェース、及び可視化システムに関する。更に、本発明は、上記方法を取り入れた解析法及びデータベースに関する。

背景技術

ヒトゲノム計画の完了に伴い、遺伝子配列、さらにはそこにコードされたタンパク質配列情報が網羅的に集積されてきている。現在、新しい診断方法や新薬の創出を目指して、これらの配列情報やタンパク質を用いた機能解析が、盛んに行われている。タンパク質の機能を調べる上で、タンパク質－タンパク質相互作用を知ることには、非常に重要な意味がある。それは、他の生体内物質との相互作用が、タンパク質の機能そのものであるからである。タンパク質－タンパク質相互作用以外にも、遺伝子のライブラリー毎の発現プロファイルやタンパク質－低分子化合物相互作用などのように、二つの物質、一般化していえば二つの事象間の相関関係情報は、生体内物質のシステムとしての機能の解明に寄与すると考えられる。タンパク質－低分子化合物相互作用に関して言えば、この相互作用データは、低分子化合物がどのようなタンパク質群に影響を与えるのか、また逆にタンパク質はどのような低分子化合物によって影響を与えられるかについての知見を提供する。タンパク質の発現量や発現タイミングに関する情報や、タンパク質と他のタンパク質との相互作用情報があつたとき、これらの情報とタンパク質－低分子化合物相互作用情報とを組み合わせることで、タンパク質の生体内の機能がわかり、その機能が低分子化合物によってどの

BEST AVAILABLE COPY

ように変化するかを予測することができる。言い換えれば、低分子化合物が医薬品になりうるかどうか予測することができるということである。このような背景に基づき、近年、さまざまな二つの生体関連事象間で大規模なデータ収集が行われ始めている。そこでは、データ量が増大すればするほど、データ全体を概観しそこから特徴を抽出することが困難になるという問題があった。また、データ量が増大すれば、個別データの詳細な参照が多数必要となって、個別サイトの観察が頻繁になるという問題もあった。そこで、大量の相関データから、その中に埋もれている情報を有効に抽出するために、情報可視化方法の重要性が増大している。

大量の相関データの可視化方法として、一方の事象を行、もう一方の事象を列にとった行列を考え、この行列の交差するセル内に二つの事象間の相関データを記述する表示方法がある。発現プロファイルでは、行列のセル内に発現強度に応じた色彩を表示する方法が、一般に用いられている。タンパク質－タンパク質相互作用の可視化においても、行列のセル内に相互作用に応じた色彩あるいは濃淡を表示する方法が行われている。タンパク質－低分子化合物相互作用の可視化においても、行列のセル内に相互作用に応じた「++」、「+」などの定性的情報を表示する方法が行われている（特許（PCT:WO 02/23199 A2））。

二つの事象間の相関関係情報を行列で表示する方法においては、行列上の相関データのパターンをもとにクラスタリングを行うことが一般的に行われている。得られたクラスター内の事象がどういう事象かを解析することによって、相関情報と各事象の特徴間の関連がわかる。また同様に、各事象の特徴によって事象のソートを行い、得られた相関情報パターンと事象の特徴を比較することによって、相関情報と各事象の特徴間の関連がわかる。このように、行列による相関データの可視化方法においては、相関情報のパターンと各事象の特徴を両方観察できることが重要である。

従って、情報閲覧の有効な方法として、まず、データ数の規模が大きい相関データに対する行列表示を行い、相関データパターンによるクラスタリングや各事象の特徴による事象のソート等によって特徴的なパターンを

同定する。その後、同定したパターンの構成要素に関する特徴量や相互作用情報の詳細情報にアクセスしていくことによって、得られたパターンの持つ意味について考察していくことが可能になる。さらに、上述したクラスタリングやソートと異なる方式でクラスタリングやソートを再度行い、得られた相関データパターンの全体を観察し、その中で先に注目した個別の相互作用と事象がどのようなクラスターに属しているかを調べることによって、新たな発見に繋がる可能性が生じる。このように、大量の相関データ行列表示と個別の相関データ表示との間で交互に行き来を繰り返すことによって、相関データに関する新しい知識の発見が可能になると考えられる。

しかしながら、従来の行列による相関データの可視化方法においては、データ数の規模が大きく変動した際に、規模に応じた適切な情報が得られないという問題があった。例えば、画面の画素数が縦横 1,000 ピクセル×1,000 ピクセル程度（大きさで言えば 30cm×30cm）程度であるとしよう。データ規模が数十個～百個のオーダーの場合は、一つのセルあたりの画素数は 10～数十ピクセル×10～数十ピクセルで、大きさにして数 mm²～1cm²程度になり、色彩あるいは濃淡のパターンと各データポイント一つ一つが同時に観察可能である。

しかし、データ規模が数百個以上に増大した場合は、一つのセルあたりの画素数は数ピクセル×数ピクセル以下で、一つのセルの大きさは 1mm²以下程度になる。この場合は、セルが小さすぎてパターン情報が複雑になると同時に、セル一個一個の認識が困難になってくる。また、描画時間がかかるという問題も生じてくる。このように、データ規模が数百個以上に増大した場合には、一定数のセルあるいはクラスターに対応した複数のセルをまとめて一つの相関データを記載するパターンの粗視化を選択することで、一つのセルのサイズが数 mm～1cm×数 mm～1cm 程度になり、相関データパターンと各データポイント一つ一つが同時に観察可能になる。従来は、この操作をユーザーがマニュアルによって実施する必要があり、手間がかかっていた。

逆に、行ないし列の規模が数十個以下に減少した場合は、一つのセルあたりの画素数が数十ピクセル×数十ピクセル以上で、一つのセルの大きさにして数 cm^2 以上と大きいにも関わらず、セル当りの情報量が色彩で表現される程度の情報量のままであるため、画面全体から得られる情報量が減少してくる。画面全体から得られる情報量を増加させるために、個々のセルに関する情報を参照しようとするれば、個々のセル毎に別の情報ソースにアクセスする必要性が生じてくる。この場合、相関データパターンと、パターンを構成する複数のセルに関する情報を同時に参照することが困難であり、また手間も大きかった。

発明の開示

本発明が解決しようとする課題は、二つの事象間の相関データを行列形式で表示する可視化方法において、相関データパターンとパターンを構成する複数のセルに関する情報を、データ数の規模の変動に応じて適切な形式で、同時に観察する手段を提供することである。

背景技術で述べてきたように、二つの事象間の相関データを行列形式で表示する可視化方法において、相関データパターンとパターンを構成する複数のセルに関する情報を同時に観察するためには、相関データ規模の大小によって、相関データパターンの粗視化（クラスタリング等によって複数のセルをまとめて要約する作業）や、セル毎の情報の他ソースへのアクセス等の作業を実施する必要があった。しかも、従来の方法では、これらの作業はマニュアルによって行わなければならなかった。従来技術で述べたように、大量の相関データから有効な知識を発見するためには、相関データの全体としての観察と少数データの詳細な観察を交互に繰り返す作業が必要である。従来マニュアルによる方法は、この繰り返し作業を行う際の効率が非常に低かった。したがって、多量のデータから創薬に有用な知識を抽出する効率も低かった。

前記課題を解決するため、本発明による二つの事象間の相関データを行列形式で表示する画面表示システムは、データ数の規模の変動に応じて、

予め用意された複数の単位相関データあたりのデータの集積度が異なるデータ表示形式の中から一つを自動的に選択し、また、個々のセルに関する情報（相関や各事象に関する情報）について予め用意された複数の要約度が異なる表示方法の中から一つを自動的に選択して、相関データと個々のセルに関する情報を表示することを特徴とする。

二つの事象間の相関データの典型例としては、一方の事象は蛋白質、もう一方の事象は低分子化合物、事象間の相関データは蛋白質－低分子化合物間の相互作用の強さである。また、両方の事象共に蛋白質で、事象間の相関データは蛋白質－蛋白質間の相互作用の強さ、あるいは蛋白質間の配列類似性であってもよい。さらに、一方の事象は遺伝子、もう一方の事象は遺伝子が由来する cDNA ライブラリーであって事象間の相関データは遺伝子の cDNA ライブラリー毎の発現強度であってもよい。また、両方の事象共に低分子化合物で、事象間の相関データは低分子化合物間の構造類似性や薬効上又は副作用上の相互作用であってもよい。

多量の相関データ、例えば蛋白質－低分子化合物相互作用データ、から有用な知見を取り出すに当たっての解析は二つのステップによって行われる。第一のステップは、データの並べ替えである。並べ替えの方法は複数ある。データは蛋白質の物性のうちの 1 つに対して昇順、または降順に並べ替えることができる。また、蛋白質のある 1 つの分類ごとにまとめて並べ替えることもできる。同様に、データは化合物の物性のうちの 1 つに対して昇順、または降順に並べ替えることができる。また、化合物のある 1 つの分類ごとにまとめて並べ替えることもできる。さらに、蛋白質－低分子化合物の相互作用強度に類似性に基づいて、類似した相互作用を持つ蛋白質や低分子化合物が互いに隣にくるように、蛋白質や低分子化合物を並べ替えることもできる。相互作用強度に基づく、蛋白質間および低分子化合物間の類似度の計算はクラスタリングと呼ばれるものであり、特に 2 つの事象間の相互作用情報から知識を抽出するにあたり有用なデータ分類、並べ替えの手法である。クラスタリングにより、相互作用強度を表示する表は、強度が強い部分と、弱い部分が分離した形で表示されるようになり、

強度が強い部分を濃く色づけして表示すれば、その部分は海に浮かんだ島に見立てることができる。一つ一つの「島」をクラスターと呼ぶ。クラスターは強度の強い部分ものほど、着目度が高いので、クラスタリング結果は、一つ一つのクラスターを強度における降順に対角線上に並べることで、重要なクラスターから順番に詳細な観察ができる。

さて、第二ステップは、クラスタリングの結果得られたこれらクラスター一つ一つの詳細な解析である。まず、クラスターを形によって以下の3つに分類する。長いクラスター、大きいクラスター、シングルトンの3つである。長いクラスターは、一つの低分子化合物に対して複数の蛋白質が強く相互作用するときや、一つの蛋白質に対して複数の低分子化合物が強く相互作用するときに形成されるクラスターである。大きいクラスターは、複数の低分子化合物と複数の蛋白質の組み合わせの中の全部または一部が互いに強く結合するときに形成されるクラスターである。最後にシングルトンは、一つの低分子化合物と一つの蛋白質の組み合わせにおいて、特異的に強い相互作用が見られた場合に形成されるクラスターである。

上記の3種類のクラスターごとに異なる解析を行う。まず、長いクラスターの解析では、複数ある低分子化合物（又は蛋白質）の共通部分を抽出する。共通部分とは数値で表される物性の取りうる範囲であってもよいし、構造上の類似した特徴であってもよい。また、化合物、あるいはタンパク質の属性が複数の要素からなるプロファイルで表現される場合であってもよい。これら共通部分は対象蛋白質（又は対象低分子化合物）との結合を生み出すために必須のファクターであると考えられる。特に、対象蛋白質との結合に関与する低分子化合物の構造上の特徴部分はファーマコフォアと呼ばれる概念に通じ、ファーマコフォアは創薬の上で重要な役割を持つ情報である。逆に、対象低分子化合物との結合に関与する蛋白質の構造上の特徴部分は蛋白質の「結合ポケット」、「へこみ」などの言葉で表現される活性部位であり、これらの活性部位の形状を詳細に観察することで、低分子化合物の構造修飾により、クラスター内のある蛋白質に対する相互作用は保持するか、クラスター内の別の蛋白質に対する相互作用が失われる

ような分子のデザインも可能となる。共通部分構造の抽出ができたとしたら、次にクラスターに属さない低分子化合物（又は蛋白質）のうち、同様の共通部分構造を持つものを検索する。検索の結果得られた低分子化合物（又は蛋白質）はクラスターの定義により対象蛋白質（または対象低分子化合物）との強い相互作用が認められなかったものである。したがって、クラスターに属する低分子化合物（又は蛋白質）とクラスターに属さないが同様の共通構造をもつ低分子化合物（又は蛋白質）を明快に区別するような物性や構造上の特徴の抽出も重要である。長いクラスターがあるとき、それらひとつひとつの要素における相互作用強度は異なると考えられるが、要素をクラスター内で相互作用強度の順序に並べ替えたときに、相互作用強度の変化を説明できるような物性や構造上の特徴の抽出は、低分子化合物であれば、それらを最適化してより対象蛋白質に特異的に結合する低分子化合物のデザインにつながる有用な知見となる。

大きなクラスターの解析では、基本的には長いクラスターの解析を蛋白質方向と低分子化合物方向に複数回繰り返す。大きなクラスターの解析では、長いクラスターの解析の結果得られるような知見の複数倍の知見が得られるので、それらを統合することで、より確実な低分子化合物や蛋白質の物性上、構造上の特徴を明らかにすることができる。

化合物、あるいはタンパク質の属性が複数の要素からなるプロファイルで表現される場合として、タンパク質との相互作用プロファイル、タンパク質の発現プロファイル、低分子化合物の薬効や副作用プロファイル等が考えられる。これらのプロファイルを用いれば、蛋白質-低分子化合物相互作用から得られたクラスター中のタンパク質や低分子化合物を、これらのプロファイルでみた場合の共通性によって、分類することが可能である。

最後に、シングルトンの解析であるが、ここでは、長いクラスターや大きなクラスターの解析に用いたような共通部分構造の抽出というアイデアを用いることはできない。しかしながら、シングルトンの構成要素である低分子化合物と蛋白質は特異的に結合するペアであることから、このペアの生物学的重要性を考察することがもっとも重要である。このペアは医

薬品とそのターゲット蛋白質の関係にあるかもしれないし、副作用を引き起こす低分子化合物とそのターゲット蛋白質の関係にあるかもしれないし、結合しても生物学的にはあまり意味のある変化を引き起こさないかもしれない。もし、このペアが医薬品とそのターゲット蛋白質の関係であれば、化学修飾によってより特異的に対象蛋白質に結合する低分子化合物がデザインできる可能性がある。

最後に第二ステップでの、クラスターの解析結果をデータベース化する。上記で示した相互作用クラスターに共通の属性の解析結果、及文献や特許から抽出されたび関連する既知情報（蛋白質－蛋白質相互作用の情報、低分子化合物と蛋白質の複合体情報、低分子化合物の毒性情報、蛋白質の発現情報など）を収集し、データベース化を行う。このデータベースには、クラスター解析結果からの既知関連情報の検索や、既知情報からのクラスター解析結果の検索機能を搭載する。この検索機能を活用することにより、ユーザーは、相互作用クラスターの分子生物学的あるいは薬学的な解釈を行うことが可能になる。

以上の二ステップの解析によって、多量のデータから、創薬に有用な知見を抽出することを目指しているが、第一ステップではデータがあまりに多量であり、表の形ですべてのデータを表示し、そこからデータの意味を理解するのは困難であるという問題がある。逆に第二ステップでは、クラスターごとにデータを詳細に観察するため、より詳しいデータが画面上で見られなくはいけない。実際の解析においては、これらのステップの繰り返しによってデータ解析を進めていくため、多量のデータの簡潔な表示と、比較的少量のデータの詳細な観察が容易に繰り返されるようなシステムが求められている。

本発明による画面表示方法は、データ表示形式として、(A) 関連データの要素そのもの、例えば低分子化合物とタンパク質の結合定数、を画面表示データ単位とする表示形式（個別データ表示形式と呼ぶ）、(B) 複数の相互作用データのまとまりを画面表示データ単位とする表示形式（関連データのパターンや事象の特徴に基づくクラスタリングから得られたクラス

ターを、複数の相互作用データのまとまりとする。そこで、クラスター表示形式と呼ぶ)、(C) 複数の相関データの統計値を画面表示データ単位とする表示形式(統計表示形式と呼ぶ)の三つを有することを特徴とする。相関データの統計値とは、クラスターの数そのものや、クラスターの各要素について別のデータソースから得られる関連情報の数などをいう。

本発明による画面表示方法は、個々のセルに関する情報(相関や各事象に関する情報)の表示方法として、情報量に依存して設定された複数の要約度に従った表示方法を有することを特徴とする。要約度は、一つの事象を表現する際の情報量が小さいほど高い値として定義される。

本発明によって定義される複数の要約度は、以下のとおりである。データフィールドに格納されている意味的に重複しない全ての情報を画面に出力するとき、データは要約されていないので、データの要約度は0であるとする。異なる種類のデータフィールドに対して、それぞれ複数の要約度に対応するデータのフォーマットを定義しておく。例えば、指数部分を含む実数データの表示において、

要約度0ではフィールド値そのものを表示、

要約度1では指数部分のみを表示、

要約度2では指数部分の値を五つのクラスターに分類し、クラスターに対応する色で情報を表示、

要約度3ではある閾値以上のもののみ色をつけて表示、

とすることができる。また、階層構造を表している文字列データの表示において、

要約度0では階層構造のそれぞれの定義を階段状に表示、

要約度1では階層構造の最上層または最下層の定義のみを表示、

要約度2では階層構造の最上層または最下層に対応する情報をシンボルや色彩に射影して表示、

要約度3では階層構造の最上層の値に対応する色をつけて表示、

とすることができる。

本発明による画面表示方法は、データ数の規模の変動に応じて、上述し

た複数のデータ表示形式の中から一つを自動的に又は手動で選択するステップ、また上述した個々のセルに関する情報（相関や各事象に関する情報）の要約度の異なる複数の表示方法の中から一つを自動的に、あるいは手動で選択するステップ、及び選択したデータ表示形式と要約度を用いて、相関データと各事象に関する情報を表示するステップ、とを含むことを特徴とする。

本発明によるデータ表示形式と要約度を自動的に選択する場合、画面に表示される情報量をユーザーが認識可能な最大の情報量付近の一定の値の近傍に留めるような選択を行うことを特徴とする。別の言い方をすれば、一つの画面に関連するすべての情報が表示されることを基準としてデータ表示形式と要約度を自動的に選択する。ただし、画面の少々スクロールを許してよい。

以上のことを行うことによって、二つの事象間の相関データを行列形式で表示する可視化方法において、相関データ規模の大小によって、相関データパターンの粗視化や、セル毎の情報の他ソースへのアクセス等の作業をマニュアルで実施することなく、相関データパターンとパターンを構成する複数のセルに関する情報を、データ数の規模の変動に応じて自動的に選択された適切な形式で、同時に観察することが可能になる。これによって、相関データの全体としての観察と少数データの詳細な観察を交互に繰り返す作業を、従来のマニュアル操作に比べ大幅に効率的に実施することが可能になり、大量の相関データからの有効な知識の発見を効率的に行うことが可能になる。

図面の簡単な説明

第1図は、データ可視化のフローチャートである。第2図は、低分子化合物とタンパク質の相互作用データの画面表示例である。第3図は、相互作用データプロファイルを用いたクラスタリング結果に基づいてソートされたデータの画面表示例である。第4図は、行および列の特徴量を用いたクラスタリング結果に基づいてソートされたデータの画面表示例である。

第5図は、クラスター表示形式での情報表示例である。第6図は、個々データ表示形式での4つの要約度別の情報の画面表示例である。第7図は、データの表示形式とデータの要約度を決定するためのルールである。第8図は、低分子化合物物性テーブルについての要約ルール決定表である。第9図は、関連情報抽出方法の概要である。第10図は、関連情報の抽出結果である。第11図は、本発明を実装したユーザーインターフェースの画面例である。第12図は、PLDデータを低分子化合物を25、タンパク質を15のグループに分割するようなクラスターにした前後の結果である。第13図は、PLDデータのクラスタリング結果の2種類の表示例である。第14図は、低分子化合物タンパク質間相互作用の行列、及び隣接して表示されたタンパク質の細胞組織における発現プロファイル行列と低分子化合物の有害事象行列である。第15図は、実験によって得られた低分子化合物タンパク質間相互作用情報と文献等から得られた既知低分子化合物タンパク質間相互作用情報を、一つの行列内に同時に表示した例である。第16図は、2つの事象間の相互作用として、医薬品低分子化合物の化学構造類似度情報と有害事象行列による分類情報を、一つの行列内に同時に表示した行列である。第17図は、蛋白質と低分子化合物の複合体情報を二次元の表を使って表示した例である。

以下、各図で用いられている符号を説明する。

101：ユーザー操作、102：内部計算、103：データ処理、104：蛋白質－低分子化合物相互作用データベース、105：各種相関関係テーブル、106：表示データ、107：データ表示形式と要約度決定ルール。
201：低分子化合物のラベル、202：タンパク質のラベル、203：マトリクス部分、204：分子量、205：アルファヘリックスとベータストランドの数、206：相同性に基づくクラスタリング情報。
301：低分子化合物クラスターA、302：低分子化合物クラスターB、303：低分子化合物クラスターC、304：タンパク質クラスターA、305：タンパク質クラスターB、306：タンパク質クラスターC、307：特定の低分子化合物とタンパク質の組からなるクラスター、308：一つ

のタンパク質について特異的に相互作用をもつ化合物の組からなるクラスター。

401：分子量の比較的大きなクラスターA、402：中程度の分子量を持つクラスターB、403：分子量の比較的小さなクラスターC、404：アミノ酸配列の相同性に基づいてクラスター1、405：アミノ酸配列の相同性に基づいてクラスター2、406：比較的相互作用が高い領域。

501：ラベル、502：クラスターに属する要素の数、503：クラスターに属する要素のリスト、504：マトリクス部分。

601：要約度0における画面表示、602：要約度1における画面表示、603：要約度2における画面表示、604：要約度3における画面表示。

701：要約度、702：データ項目、703：場所、704：要約ルール、705：ルール「そのまま」、706：ルール「色彩(200, 300, 400, 500)」。

801：条件、802：表示形式、803：要約度。

901：タンパク質－低分子化合物相互作用テーブル、902：タンパク質－タンパク質相互作用テーブル、

903：タンパク質－発現テーブル、904：低分子化合物－低分子化合物相互作用テーブル。

1101：表示モードの変更ボタン、1102：要約度の変更ボタン、1103：関連情報取得ボタン、1104：アクションに関連する機能群、1105：選択に関連する機能群、1106：関連情報表示画面。

1201：クラスタリング前の行列、1202：クラスタリング後の行列、1203：クラスタリングの結果に意味を見出せる領域、1204：クラスタリングの結果に非類似な相互作用データが入り混じった領域

1301：クラスターを単位とした行列データの一部を要約度2で画面表示した例、1302：クラスターに属する低分子化合物の数、1303：クラスターに属するタンパク質の数、1304：クラスターに属する相互作用の数、1305：個々のタンパク質と低分子化合物を単位にした行列による表示、1306：縦12×横1の行列で表されるクラスター、13

07 : クラスターの要素である化合物群の物性値、1308 : 化合物の物性と相互作用強度が対応するクラスター、1309 : クラスター1308の要素である化合物の物性、1310 : クラスター1308の相互作用強度と化合物の物性値1309の値を3段階の値に射影した表

1401 : 低分子化合物タンパク質間相互作用の行列、1402 : 細胞組織における発現プロファイル行列、1403 : 有害事象行列、1404 : 低分子化合物タンパク質間相互作用クラスター、1405 : 低分子化合物タンパク質間相互作用クラスター、1406 : 低分子化合物タンパク質間相互作用クラスター領域、1407 : 低分子化合物タンパク質間相互作用クラスター領域、1408 : 低分子化合物タンパク質間相互作用クラスター領域、1409 : 低分子化合物タンパク質間相互作用クラスター領域、1410 : 細胞組織における発現プロファイル、1411 : 細胞組織における発現プロファイル、1412 : 有害事象行列のプロファイル、1413 : 有害事象行列のプロファイル

1501 : 低分子化合物タンパク質間相互作用行列、1502 : 既知相互作用情報に基づくクラスタリングによって得られたクラスター、1503 : 既知相互作用情報のクラスターに属さない、実験によって得られた相互作用

1601 : 医薬品低分子化合物の化学構造類似度情報と有害事象行列による分類情報を同時に表示した行列、1602 : 化学構造類似度情報に基づくクラスタリングを行い、得られたクラスター、1603 : 低分子化合物C5、C4間ペア、1604 : 化学構造類似度がない化合物ペア

1701 : 蛋白質と低分子化合物の複合体の重心間距離情報を表示した行列、1702 : 低分子化合物を含むクラスター、1703 : タンパク質-低分子化合物複合体のモデル

発明を実施するための最良の形態

以下、図面を参照して本発明の実施の形態を説明する。

[実施例1]

二つの事象間の相関関係として、蛋白質、低分子化合物、DNA等の生体内物質間相互作用を考える。着目する二つの事象として「低分子化合物」と「タンパク質」間の相互作用データを扱う場合の実施例を、以下に説明する。ここで、相互作用データとは、Protein Data Bank (PDB, <http://www.pdb.org>)中に低分子化合物とタンパク質の複合体データがあるか、ないかという情報や、実験的に低分子化合物とタンパク質との間の結合の度合いを測定したデータである。タンパク質の特徴データとしては、各種外部データベースの情報や計算されたクラスタリング結果を持つ。例えば、SWISSPROT(<http://www.expasy.ch/sprot>)のIDや、アミノ酸配列相同性に基いたクラスタリング結果、Gene Ontology(<http://www.geneontology.org>)に基づくアノテーション情報、溶媒への溶解度などである。低分子化合物の特徴データとしては、分子名、分子量、薬効分類、その他、電荷分布や親水・疎水性、立体構造、水素結合のドナー・アクセプター数、官能基の種類や数など様々な分子特性値を持つ。

まず、第1図を用いてデータ可視化のフローチャートを説明する。ユーザー操作101はデータと実行するアクションを選択する部分である。アクションには、データ取得102とデータ処理103がある。データ取得には、各種検索条件による蛋白質-低分子化合物相互作用データベース104からの検索によるデータ取得、表示画面上で指定された蛋白、あるいは低分子化合物に関連した各種相関関係テーブル105からのデータ取得がある。データ処理には、表示画面上で指定されたエントリーに対するクラスタリング等の処理や表示スケールの変更等の処理がある。取得、あるいは処理されたデータは表示データ106として扱われる。次に、表示データに対して、データの表示形式と要約度が決定される。データの表示形式と要約度は、表示データのデータ数に応じて、予め用意されたデータの表示形式と要約度決定ルール107に基づいて決定される。決定されたデータの表示形式と要約度に従い、データの画面表示108が行われる。各種相関関係テーブルとしては、タンパク質-タンパク質相互作用テーブル、

タンパク質の発現プロファイルテーブル、低分子化合物－低分子化合物間の構造類似性や、薬効上または毒性上の相互作用テーブル等が考えられる。

本発明の要点である、「データの表示形式と要約度が、表示データのデータ数に応じて、予め用意されたデータの表示形式と要約度決定ルールに基づいて決定される」という点について、以下詳細に説明する。

まず、データの表示形式について説明する。第2図に低分子化合物とタンパク質の相互作用データの画面表示例を示す。行列の縦方向に低分子化合物のラベル201、横方向にタンパク質のラベル202を並べ、行列部分203には実験的に測定されたタンパク質と低分子化合物の間の結合定数のうちある閾値より上のものに関して結合の強さ別に色の濃さを変えて表示している。また、化合物ラベルの左側には化合物の特徴量として分子量204を表示し、タンパク質ラベルの上側にはタンパク質の特徴量として、アルファヘリックスとベータストランドの数205と蛋白質相互の相同性に基づくクラスタリング情報206を表示している。

表形式で画面表示された相互作用データについては、相互作用データプロファイルに基づくクラスタリング、あるいは、タンパク質の特徴量や、低分子化合物の特徴量に基づくクラスタリングを行い、得られたクラスタリング情報に基づいてデータを並べ替えて表示することが可能である。

相互作用データを用いたクラスタリングは、例えば以下の方法によって行う。ひとつの低分子化合物 C_i に着目して、それと各タンパク質 P_j の相互作用強度プロファイル I_{ij} ($j=1, \dots, N_p$, N_p はタンパク質数) を考える。次に、全ての低分子化合物間で総当りの相互作用強度プロファイル間距離を計算する。低分子化合物 C_i と低分子化合物 C_k 間の相互作用強度プロファイル間距離 D_{ik} は、低分子化合物 C_i とタンパク質 P_j 間の相互作用強度が I_{ij} とすれば、例えば以下の式によって計算される。

$$D_{ik} = \sqrt{\sum (I_{ij} - I_{kj})^2}$$

上式中の和は $j=1, \dots, N_p$ についてとる。

この式によって得られた総当りの D_{ik} に対して閾値を設けることによって、低分子化合物をクラスタリングすることが可能である。次に、ひとつのタンパク質 P_i に着目して、それと各低分子化合物 C_j の相互作用強度プロファイル I_{ij} ($j=1, \dots, N_c$, N_c は低分子化合物数) を考える。低分子化合物の場合と同様に、全てのタンパク質間で総当りの相互作用強度プロファイル間距離を計算することによって、タンパク質をクラスタリングすることが可能である。

上記のクラスタリングを実際に行った結果が、第3図に示されている。

低分子化合物は3つ、タンパク質も3つのクラスターに分類され、その結果は低分子化合物のラベル上に低分子化合物クラスター A3 0 1、低分子化合物クラスターB3 0 2、低分子化合物クラスターC3 0 3として、またタンパク質のラベル上にタンパク質クラスター A3 0 4、タンパク質クラスターB3 0 5、タンパク質クラスターC3 0 6として色の濃さで識別表示されている。クラスター毎に相互作用データである結合定数の平均値が内部で計算され、クラスターは結合定数の平均によって上から下、左から右へ降順にソートされている。したがって、全体的な傾向として、マトリクス部分の左上のほうに結合定数の高い（色の濃い）セルが集まり、右下のほうには結合定数が低い又は閾値以下の結合しかないセルが集まっている。このような相互作用プロファイルに基づいたクラスタリングを行うことによって、特定の低分子化合物とタンパク質の組からなるクラスター3 0 7や、一つのタンパク質について特異的に相互作用をもつ多くの化合物を含むクラスター3 0 8などが視覚的に明らかになる。創薬研究への応用として、相互作用プロファイルに基づいて作成された低分子化合物のクラスターに共通する母核構造を抽出して、それを薬物の機能を担うファーマコフォアとして構造展開の種とするアプローチが可能である。

同様に、分子量をいくつかの区分に分けてクラスタリングしたり、タンパク質のアルファヘリックスとベータストランドの数をあるルールに従って分類したりすることが可能である。そして、分子量に基づくクラスター、

アルファヘリックスとベータストランドの数に基づくクラスター、或いはあらかじめ計算されているアミノ酸配列の相同性に基づくクラスターのそれぞれについて表示データを並べ替えることができる。特に、ある特徴量についてデータを並べ替えた結果、特徴的な結合定数の色彩パターンが表れた場合には、その特徴量と結合定数が密接に関連していることを知ることができる。

第4図に、データを低分子化合物側については分子量、タンパク質側についてはアミノ酸の相同性にもとづいてクラスタリングをし、クラスタリング結果によって表を並べ替えた結果を示す。低分子化合物は分子量によって分子量の比較的大きなクラスターA401、中程度の分子量を持つクラスターB402、分子量の比較的小きなクラスターC403に分類されており、データ全体は分子量について降順にソートされている。タンパク質は、アミノ酸配列の相同性に基づいてクラスター1、404とクラスター2、405が画面上に示されている。ここでは、クラスターBに属する低分子化合物が相互作用マトリクスの中では比較的高い相互作用領域406と重なっているように見える。一方、アミノ酸の相同性に基づくクラスタリング結果と相互作用強度の間には明白に視認できるような相関は見当たらないようである。このように特徴量に関してクラスタリングを行い、その結果によってデータを並べ替えることによって、相互作用データをよく説明するような特徴量を発見できる可能性がある。低分子医薬品がもつ特徴量（分子特性）としてよく知られているものに Christopher A. Lipinski 博士の“Rule of five” (Advanced Drug Delivery Reviews, 23 (1997) 3-25) があるが、特徴量によるクラスタリング結果と相互作用データを同時に可視化することで、特定の実験データを説明する特徴量や、特定のタンパク質の標的となりうる低分子化合物が持つべき特徴量をルール化することも可能であると考えられる。

第3図あるいは第4図の表形式のデータ表示においては、表の個々のセルが一つのタンパク質と低分子化合物の相互作用に対応している。これをここでは「個々データ表示形式」と呼ぶ。しかし、個々データ表示形式に

においてはタンパク質の数や低分子化合物の数が増えるにしたがって、表のサイズが大きくなり、データ全体の把握が難しくなってくるという欠点がある。すなわち、データ数の増大に応じて表の個々のセルのサイズを変えなければ、表全体が画面に入りなくなり、データ全体の様子を一望することができなくなる。逆に、表の個々のセルのサイズを小さくすることによって、表全体を画面内に収めるようにすると、セルに表示された相互作用データのパターンが細くなり、その特徴の認識が困難になる。そこで、データ数が増大した場合も一望して表全体の相互作用パターンを認識可能にするために、第3図あるいは第4図における個々のクラスターを表上の一つのセルとして情報を表示することを可能にした。これをここでは「クラスター表示形式」と呼ぶ。

第5図において、クラスター表示形式での情報表示例を示す。ラベル501にはクラスターの番号が入り、特徴量としてはクラスターに属する要素の数502と、クラスターに属する要素のリスト503が示されている。マトリクス部分504にはクラスターごとの測定データの平均値が色の濃さによって表示され、クラスターを構成する要素の数が数値によって示されている。個々データ表示形式による情報表示とクラスター表示形式による情報表示の切り替えが可能である。また、一つの表示形式における行や列の並べ替え、削除などの操作はもう一つの表示形式に反映される。クラスター表示形式においては、似たタンパク質同士、似た低分子化合物同士がクラスターを形成することから、代表的なデータを取りこぼすことなく可視化することができる。それと同時にクラスターの数調節することによって、相互作用データの数が多いたときも表示される表の行数、列数をコントロールできる。

個々データ表示形式とクラスター表示形式に相補的な情報表示形式として、「統計量表示形式」がある。これはデータの全部または一部に対して平均値、標準偏差などの統計計算を行い表示したり、異なるデータソースから抽出されたデータの件数を表示したりする形式である。統計量表示形式においては、相互作用データの数にかかわらず、データの全体像を把握

することができる。特に、データ数が増大した場合には、クラスター表示形式においても、一望して表全体の相互作用パターンを認識することが困難になってくる。このような場合に、統計量表示形式は、データの全体像を把握するという観点で非常に有効である。

本発明においては、表示形式を複数用意すると同時に、行列の各セル中に表示する情報として、要約の程度を変えたものを複数用意しておき、その中からデータ数に応じたものを選択して用いることを特徴としている。

タンパク質と低分子化合物の相互作用データの表示においては、4つの要約度（0 - 4）を用意する。要約度0では、データベースに格納されている情報や、そこから計算された統計量などをもれなく表示する。要約度1では、一つのセル当たり64文字までの文字データ、記号、色彩を表示できる。データベース中のテキストフィールドで64文字以下のものや、たとえ長いものであっても64文字以下に情報を削減できるものであれば表示可能である。要約度2では、一つのセル当たり8文字までの文字データ、記号、色彩を表示できる。要約度3では、文字データは表示しない。全ての情報を色彩で表現する。

実装においては、要約度0における情報表示はフリーフォーマットとし、要約度1では一つのセルのサイズを縦60ピクセル×横120ピクセルとして、その中に16文字×4行分のテキストを表示する領域を確保する。要約度2では一つのセルのサイズを縦20ピクセル×横60ピクセルとして、その中に8文字×1行分のテキストを表示する領域を確保する。要約度3では一つのセルのサイズを縦5ピクセル×横5ピクセルとした。原理的には一つのセルのサイズを最低1ピクセル×1ピクセルにまで縮小することは可能であるが、マウスを使って個々のデータを操作可能なセルサイズを選択している。

これら4つの要約度における画面表示は、切り替え表示が可能である。第6図に個々データ表示形式での4つの要約度別の情報の画面表示例を示す。

要約度0における画面表示601では、相互作用のデータ、低分子化合

物のデータ、タンパク質のデータが詳細に表示されている。表示フォーマットは自由であり、タンパク質や低分子化合物の構造なども表示し操作することが可能である。

要約度 1 における画面表示 6 0 2 では、タンパク質関連の各種外部データベースへアクセスするためのキー、低分子化合物の名前や薬効、また相互作用の測定データの詳細な数値などを表示している。

要約度 2 における画面表示 6 0 3 では、表示される文字データは 8 文字までに限られるので、行や列を識別するためのラベルや、相互作用の測定データの主要な値などの限られた情報を表示している。

要約度 3 における画面表示 6 0 4 では、各セルがとる値を色彩情報に変換して表示している。これによって類似したデータを色彩のパターンから視認することができる。

選択されたデータ項目について、要約度によってどのように情報を要約するのかに関してルールを作る必要がある。基本的なルールは、要約度 0 においては、すべての情報の表示、要約度 1 と 2 においては文字の長さに応じた情報表示、要約度 3 においては色彩表示となっている。この基本的なルールにのっとり、詳細な要約のルールを、データベースに存在するそれぞれのデータ項目について定義する必要がある。

第 7 図に、一例として、低分子化合物特徴テーブルについての要約ルール決定表を示す。要約度 7 0 1 に応じて、テーブル中のフィールドのうちどのデータ項目 7 0 2 を、どの場所 7 0 3 に、どのような要約ルール 7 0 4 で加工して画面表示をするかについての情報が与えられている。

フィールド名が要約ルール決定表に現れない場合は、そのフィールドは表示されないことを意味する。要約ルールが「そのまま」7 0 5 の場合、データベースに格納されているデータをそのまま表示する。別の例として「色彩 (2 0 0, 3 0 0, 4 0 0, 5 0 0)」7 0 6 の場合、値が 2 0 0 未満、2 0 0 以上 3 0 0 未満、3 0 0 以上 4 0 0 未満、4 0 0 以上 5 0 0 未満、5 0 0 以上の五つのケースについて色分け表示をする。このような要約ルール決定表をデータベース中のそれぞれのテーブルについて持つ必要

がある。

以上、3つのデータの表示形式と、4つのデータの要約度を説明した。これらを組み合わせることによって多種多様な角度からデータを可視化することが可能である。本発明は、ユーザーが見たい情報を選択すると、そのデータ数に応じて最適なデータの表示形式とデータの要約度を自動的に決定する機能に特徴がある。

データの表示形式とデータの要約度を自動決定するための入力データとして、タンパク質と低分子化合物の相互作用データの可視化の例においては、タンパク質の数 P 、低分子化合物の数 C 、タンパク質クラスターの数 P_c 、低分子化合物クラスターの数 C_c 、及び、画面上における情報表示領域のパラメータ x （高さ）、 y （幅）が必要である。クラスターの種類が複数ある場合は初期設定として登録されているクラスターの数を使用する。

第8図にデータの表示形式とデータの要約度を決定するためのルールを表形式で示す。条件801を上から順番に見ていき、条件を満たしたところで、その行に記述されている表示形式802と、要約度803を採用する。条件を満たさない場合は、次の行の条件を見る。ここで、 G 、 R 、 G_c 、 R_c は第8図中で定義された数値である。以下この表を説明する。

$P \times C$ （表示画面内のセル数に該当）が一定値（この場合は3）より小さい場合、個々データ表示で要約度0を用いる。

$P \times C > 3$ で、かつ $G \leq 11$ & $R \leq 11$ の場合は、列方向特徴量表示数と行方向特徴量表示数がともに1である場合、タンパク質の数 P 、低分子化合物の数 C 共に2以上で、9以下となる。この場合は、要約度1を用いるので、一つのセルのサイズが縦60ピクセル×横120ピクセルとなり、縦450ピクセル×横900ピクセルの情報表示領域においては、全データの表示サイズは、縦240ピクセル×横480ピクセル～縦660ピクセル×横1320ピクセルとなる。これは、情報表示領域全体の1.5×1.5倍以内のサイズである。

タンパク質の数 P 、低分子化合物の数 C が増大するに従って、図8に従い順次、要約度を2、3と大きくしていく。さらに P 、 C 数が増大した場合、

クラスター表示に切り替え、タンパク質クラスターの数 P_c と低分子化合物クラスターの数 C_c が増大するに従って、要約度を 1、2、3 と増加させていく。

以上示した表示形式と要約度の切り替えを行うための G 、 R 、 G_c 、 R_c に対する条件としては、全データの表示サイズが、情報表示領域全体の 1.5×1.5 倍以内のサイズになるような条件を設定している。データ表示領域の $n \times m$ 倍以内に全データの情報を表示するという一般化された基準を満たすようにするには、

$$x \times n \leq P \text{ (又は } P_c) \text{ and } y \times m \leq C \text{ (又は } C_c)$$

という一般化された条件を、データの表示形式と要約度の決定に用いればよい。

このようにすることによって、データの全体、あるいはその一定の倍数のデータ量を、情報表示領域内で表示することが可能になり、かつ、データ数の増減に応じて要約度を上下させることによって、セル内に、一望して認識可能でかつ最大限の情報量を表示可能になる。これにより、表示すべきデータ数にかかわらず、個別セル内から得られる情報量を最大に保ちつつ、データの全体像の観察が可能になる。

新規創薬ターゲットの発見のプロセスにおいては、タンパク質と低分子化合物の相互作用を可視化すると同時に、他の関連する生体関連の相互作用についても同時に情報を得て、包括的に情報を整理し、理解することが極めて重要である。関連する生体関連の相互作用の例として、低分子化合物同士の薬効や毒性に関する相互作用、タンパク質同士の相互作用、タンパク質と発現に関する情報などが挙げられる。本発明においては、これら関連情報を取得し、取得したデータ数に応じて、上述した表示形式と要約度の決定ルールに従い、表示することが可能である。

関連情報の取得は、以下のように行う。表示されているデータテーブル内の着目するセル領域を選択し、このセル領域に属する低分子化合物 ID とタンパク質 ID を抽出する。これらの ID を、関連データテーブル中で検索し、検索された ID に付随する情報を関連データテーブルから抽出す

る。

第 9 図に、関連情報抽出の具体的な方法を示す。タンパク質－低分子化合物相互作用テーブル 9 0 1 のうち (C5,P12) と (C9,P12) の二つに着目しているとき、タンパク質間の結合強度を 1 0 0 を最大値として規格化したタンパク質－タンパク質相互作用テーブル 9 0 2 と、発現ライブラリーにおける定性的なタンパク質の発現量を示すタンパク質－発現テーブル 9 0 3 からはタンパク質の ID が P12 であるもののうち、データが存在するものを抽出する。同様に低分子化合物間の多剤併用による効果のある・なしのデータを格納した低分子化合物－低分子化合物相互作用テーブル 9 0 4 からは ID として C5,C9 を持つもののうち、データが存在するものを抽出する。

関連情報の抽出結果は第 1 0 図のように、抽出元の表ごとに整理されて表示される。ユーザーが見たい表を選択すると、そのヒット件数に応じて自動的に情報の表示形式と要約度が設定され、設定された表示形式と要約度で情報が画面表示される。そのようにして表示された情報の一部から、また関連情報を取得することができる。したがって、本発明によって多次元の相互作用データを 1 対 1 相互作用データ間のリンクを効率的にたどることで可視化することができる。

本発明の可視化方法を実装したインターフェースにおいては、画面表示された情報のうち一部を選択し、選択されたデータに対して、複数のアクションから選択したアクションを実施し、アクションの結果得られた情報が画面表示される。第 1 1 図にユーザーインターフェースの例を示す。表示モードの変更ボタン 1 1 0 1、要約度の変更ボタン 1 1 0 2、関連情報取得ボタン 1 1 0 3 に加え、行や列の入れ替え、並べ替え、クラスタリング、削除などのアクションに関連する機能群 1 1 0 4 と、特徴的な行や列、代表的なサブセットとしての行や列などの選択に関連する機能群 1 1 0 5 を備える。また、画面上に表形式で表されているセルの一つ一つに対してマウス操作によるアクションが割り当てられていて、それによって、行や列を選択したり、関連情報表示画面 1 1 0 6 にセルの中には表示できない

長い文字列データなども表示したりできる。

[実施例 2]

本実施例では、相互作用データの並べ換えとその結果形成されるクラスターの解析結果の可視化によって、どのように創薬にとって有用な知識を抽出するかを説明する。2つの事象間の相互作用として、タンパク質と低分子化合物の間の結合強度を考える。ここで、結合強度の値は Protein-Ligand Database (<http://www.mitchell.ch.cam.ac.uk/pld/>) から取得した乖離定数で、それぞれの値は論文に収録されているものである。結合強度として、乖離定数が 10^{-5} より小さなもののみを抽出すると、相互作用情報は低分子化合物 95 種類、タンパク質 67 種類からなる行列の形で書ける。

この行列の類似性に基づいて、PLD データを低分子化合物を 25、タンパク質を 15 のグループに分割するようなクラスターにした前後の結果を第 12 図に示す。クラスタリング前の行列 1201 がクラスタリング後の行列 1202 のように並べ替えられる。クラスタリングを行う前は行列上に相互作用のあるタンパク質と低分子化合物の組み合わせを示す点が散在しているが、クラスタリングを行うことで、相互作用の強度のパターンが類似した行や列が隣接して表示される。クラスタリング結果に意味を見出せる領域 1203 では相互作用の強い領域が行列上で「島」のように浮き立って見える。しかし、クラスタリングの結果に非類似な相互作用データが入り混じった領域 1204 もあり、この領域では一つ一つの行列上の点、すなわち相互作用強度のデータは他と類似性を持たないと解釈できる。

第 13 図に PLD データのクラスタリング結果の 2 種類の表示例を示す。まず、それぞれのクラスターに属するデータは、クラスタリング結果に意味があれば、相互作用強度が類似であるはずである。そこで、クラスターに含まれる要素のすべてを 1 つの代表値で表すことにより、表の行や列の数を削減できる。代表値としては、ここでは平均値を用いた。クラスターを単位とした行列データの一部を要約度 2 で画面表示した例 1301 においてはクラスターに属する低分子化合物の数 1302 と、クラスターに属

するタンパク質の数 1 3 0 3 と、それらの積で定義されるクラスターに属する相互作用の数 1 3 0 4 が表示されている。ここでは、低分子化合物を 2 5 のクラスターにし、タンパク質を 1 5 のクラスターにしたため、表全体のサイズは 25×15 となる。タンパク質と低分子化合物の間の相互作用マトリクスの解析においては、クラスターのうち特に相互作用の強度が高い要素に着目する。したがって、1 3 0 1 に示すように、相互作用の強度順にクラスタリング結果の表を対角方向に並べ替えることはデータを着目の優先度の順序に並べ替えることに相当する。まず、 25×15 の行列の中から、最大の値が入っている要素の位置を特定する。その要素の位置が (p, q) であったとしたら、行列の 1 行目と p 行目、1 列目と q 列目を入れ替えることで、最大の値をもつ要素を行列の (1, 1) すなわち左上に移動できる。この操作を繰り返すことで、クラスタリングの結果を対角方向に並べるわけであるが、唯一の違いは、2 周目の操作においては、最大の値が入った要素を行列の 1 行目と 1 列目を除いた 25×14 の行列から探しだし、その要素を (2, 2) の位置に移動することである。また、クラスターを単位として表示された行列を、個々のタンパク質と低分子化合物を単位にした行列による表示 1 3 0 5 に戻すことができる。ここで前記のクラスターに属する相互作用の数 1 3 0 4 は 1 2 の要素を持つので、それをタンパク質と低分子化合物を単位として表示すると、縦 $12 \times$ 横 1 の行列で表されるクラスター 1 3 0 6 となる。

以下では、相互作用に基づいて得られたクラスターから、低分子化合物が持つ共通の属性を抽出する方法について説明する。上記で得られたクラスターの要素である化合物群の物性値 1 3 0 7 として、構造分類、分子量、Molar Refractivity, 水・オクタノール間の分配係数を同時にみることができる。相互作用強度におけるクラスタリング結果と物性値の同時観察から、このクラスターの要素である化合物はすべて同一の構造分類に属することがわかり、HETERO CYCLIC AROMATIC COMPOUNDS (ヘテロ環を持つ芳香族化合物) である。しかしながら、分子量、Molar Refractivity, 水・オクタノール間の分配係数といった数値情報から相互作用強度との間の関係

を説明するのは容易ではない。分子量だけを見ても200を下回るものから、900を超えるものまでである。これら多様な物性値を持つ化合物が同一のタンパク質に強く結合するという事は、これら化合物の間に、タンパク質との結合のために不可欠な部分構造があることが想像される。物性値そのものは、その不可欠な部分構造に不可欠な残りの構造が大きく違えば当然違った値になる。本発明では、化合物のラベル上をクリックすることで実際に化合物の構造を表示して見比べることができる。そのような構造の比較により化合物の共通構造や活性部位を推測することが可能である。ここでは、そのような詳細な解析は本発明の範囲外であるため割愛する。一方で、化合物の物性と相互作用強度が対応するクラスター1308も存在する。クラスター1308の要素である化合物の物性1309を観察すると、分子量、Molar Refractivity, 水・オクタノール間の分配係数のすべてにおいて、とりうる値の範囲は比較的限定されていることがわかる。Molar refractivityでいえば、8.3から11.5の間であり、log P値は2.4から4.5の間である。構造分類の面からもこのクラスターに属する化合物のほとんどは3 AND MORE RING SYSTEMS（3つ以上の環構造をもつ化合物）の分類に属する。クラスター1308の相互作用強度と化合物の物性値1309の値を3段階の値に射影した表1310の観察から物性と結合強度の間のより詳しい関係が見える。強い結合を持つための物性値の条件は、水・オクタノール間の分配係数が小であり、Molar Refractivityが中または大であることの2つを同時に満たすことである。どちらか一方を満たした場合には結合強度は中程度になり、どちらも満たさない場合には結合強度はクラスター中の化合物の中では最も弱くなる。このような例は、化合物の構造と物性を加味しながら、対応するタンパク質に対してより特異的に結合する化合物をデザインすることが可能であることを示している。本例では、Molar refractivityでいえば、9から11.5の間であり、log P値は2.4から3.3の間であるような化合物は、当該タンパク質に対してより特異的に結合する可能性があると予測される。

[実施例3]

本実施例では、相互作用に基づいて得られたクラスターから、化合物、あるいはタンパク質が持つ共通の属性を抽出する方法として、化合物、あるいはタンパク質の属性が複数の要素からなるプロファイルで表現される場合について、第14図を用いて説明する。第14図は、タンパク質の属性として細胞組織における発現プロファイル行列1402を、低分子化合物の属性として有害事象行列1403を取得し、それらを図のように低分子化合物タンパク質間相互作用の行列1401に隣接して表示したものである。タンパク質をP1～P7、細胞組織をT1～T7、低分子化合物をC1～C6、有害事象をS1～S5と表示している。ここで、タンパク質間相互作用行列は、実験によって得られたものを使用してもよいし、文献より取得したものでもよい。また、有害事象行列は、例えば、日本医薬品集DB (<http://www.japic.or.jp/publications/index3.html>) における有害事象に関する項目中に、国際医学用語集である医薬規制用語集 (MedDRA) 中の各用語が出現するか否かを調べることによって、得られる。

低分子化合物タンパク質間相互作用クラスター1404は、二つの領域1406、1407に分類可能である。これら二つの領域は、細胞組織における発現プロファイル行列においては、それぞれ異なるプロファイル1410、1411を持つ二つのタンパク質群 (P4、P5)、(P6、P7) に対応している。これにより、クラスター1404中のタンパク質は、全て共通の低分子化合物C2と相互作用するが、細胞組織における発現プロファイルにおいては、異なる二つのタンパク質群と相互作用することがわかる。このことは、この低分子化合物が医薬品の場合には、異なる生理学的機能を有する二種類のターゲットタンパク質と相互作用することを意味する。さらに相互作用する相手のタンパク質の機能を調べることにより、この医薬品の薬効との関連性について推測することが可能になると考えられる。

有害事象行列の表示からは、低分子化合物タンパク質間相互作用クラスター1405は、二つの領域1408、1409に分類可能である。これ

ら二つの領域は、有害事象においては、それぞれ異なるプロファイル 1 4 1 2、1 4 1 3 を持つ二つの低分子化合物群 (C 2、C 3)、(C 4、C 5) に対応している。これらの二つの低分子化合物群のうち、ひとつは一つのタンパク質 P 1 と相互作用するが、もう一つはもう一つのタンパク質 P 2 を加えた二つのタンパク質と相互作用することがわかる。これにより、二つのタンパク質がそれぞれ異なる有害事象プロファイルに関連していることが推測可能である。

低分子化合物、及びタンパク質の属性としての複数の要素からなるプロファイルとしては、タンパク質間相互作用、タンパク質の系統樹プロファイル、化合物の構造プロファイル (MACCS key 記述子等) 等であってもよい。これら全ての場合に、相互作用に基づいて得られたクラスターを構成する低分子化合物やタンパク質が、他の複数の要素からなるプロファイルとしての属性でみた場合に、どこがどのように異なるかを判定することが可能になる。

上述したクラスターの解析結果を、文献や特許から抽出された関連既知情報と共に格納したデータベースを構築することができる。クラスター解析結果からの既知関連情報の検索や、既知情報からのクラスター解析結果の検索機能を、本データベースに付加することによって、この検索機能を活用することにより、ユーザーは、相互作用クラスターの分子生物学的あるいは薬学的な解釈を行うことが容易に可能になる。

[実施例 4]

本実施例では、前記生体関連事象間の複数種類の相関データを、行列のセル中に同時に識別表示する方法について説明する。2つの事象間の相互作用として、タンパク質と低分子化合物の間の相互作用を考える。実験によって得られた相互作用情報と文献等から得られた既知相互作用情報を同時に表示した例を、第 15 図に示す。第 15 図では、低分子化合物タンパク質間相互作用行列 1 5 0 1 を示す。低分子化合物を、C 1 ~ C 6、タンパク質を P 1 ~ P 7 で表示する。低分子化合物タンパク質間相互作用行列の各セルを、実験と文献から得られた相互作用のそれぞれに対応した上下

二つの領域に分割し、相互作用の有無を、分割された領域に記号（実験；●、文献；○）を記載するかどうかによって表示している。図中には、文献等から得られた既知相互作用情報に基づくクラスタリングによって得られたクラスター1502を示している。クラスター1502において、実験によって得られた相互作用に着目することによって、既知相互作用情報のうち、実験によってどれだけ再現できたかを評価することが可能である。この場合、(C3、P4)のセルから、低分子化合物C3、タンパク質P4間では、文献で得られた相互作用は存在するが、実験によっては相互作用が得られなかったことがわかる。また、既知相互作用情報のクラスターに属さない、実験によって得られた相互作用1503に着目することによって、文献にはないが実験によって新しく得られた相互作用を同定することができる。

2つの事象間の相互作用として、医薬品低分子化合物の化学構造類似度情報と有害事象行列による分類情報を同時に表示した行列1601を、第16図に示す。医薬品低分子化合物の化学構造類似度情報は、例えば MACCS key 記述子 (*Reoptimization of MDL Keys for Use in Drug Discovery*, J. L. Durant, B. A. Leland, D. R. Henry, J. G. Nourse, JCICS, 2002, 42 (6), 1273-1280.) 間の類似度によって得ることができる。また、有害事象行列による分類情報は、実施例2で説明した有害事象行列における、有害事象プロファイル間の比較によって取得可能である。行列のセルを、化学構造類似度情報と有害事象行列による分類情報のそれぞれに対応した二つの領域に分割し、分割された領域への記号の記載によって、化学構造類似度情報と有害事象行列による分類情報を表示している。化学構造類似性強度を色の濃さ（●；高い類似性◎；中位の類似性△；低い類似性）によって、有害事象行列による同一クラスターへの所属の有無を○の有無によって表示している。

第16図には、化学構造類似度情報に基づくクラスタリングを行い、得られたクラスターを行列の対角線近傍に集めた結果を示している。化学構造類似度情報に基づくクラスター中の化学構造類似度と有害事象行列によ

る分類情報を比較観察することによって、化学構造類似度がどの程度あれば有害事象行列によって同一分類になるかがわかる。例えば、クラスター 1 6 0 2 において、低分子化合物 C2、C3、C4、C5 は相互に化学構造類似性が存在する。低分子化合物 C5、C4 間 1 6 0 3 では、弱い化学構造類似度が存在するが、有害事象行列によっては同一クラスターへ帰属しないことがわかる。1 6 0 4 に示すように化学構造類似度がない化合物ペアにおいて、有害事象行列によって同一クラスターになる場合は、化学構造類似度に依存しない有害事象の存在を確認することができる。

同時に表示する相関データとしては、タンパク質間の配列類似性と構造類似性、タンパク質間の配列類似性と機能の類似性、タンパク質間の配列類似性と発現プロファイルの類似性、低分子化合物間の構造類似性と薬効分類や、低分子化合物間の異なる二つの方法による構造分類、であってもよい。また、異なる実験方法によって得られた相互作用情報であってもよい。これら全ての場合に、一つの基準によって得られたクラスターが、他の基準によって得られたクラスターとどこが異なるかという情報を具体的にかつ直感的に得ることができる。

[実施例 5]

本実施例では、蛋白質と低分子化合物の複合体情報を二次元の表を使って表示する方法を説明する。二つの生体関連事象はともに蛋白質残基の C_{α} 原子および低分子化合物の重心である。ここで、蛋白質と低分子化合物は、ともに複数が複合体中に存在してもよい。これらの間の相関データとして、 C_{α} 原子間距離、低分子化合物の重心間距離、および C_{α} 原子－低分子化合物の重心間の距離を用いる。蛋白質と低分子化合物がそれぞれ一つずつの場合を、第 17 図を用いて説明する。蛋白質構造の二次元表示方法としては、蛋白質の C_{α} 原子間距離を縦、横共に残基番号順に並べた Distance Matrix Plot が長い間利用されており、本実施例における方法は、Distance Matrix Plot と類似している。しかしながら、本発明の方法では、Distance Matrix Plot のようにプロットを単に残基番号順に並べるだけでなく、 C_{α} 原子間距離、低分子化合物の重心間距離、および C_{α} 原子－低分子化合物の重心間の

距離を基に、 C_{α} 原子と低分子化合物の重心のクラスタリングを行い、クラスタのメンバーが集まるようにデータを並べ替えることが可能である。第17図には、距離情報として、一定の距離以下の場合にセル中に●を記載し、さらにクラスタリング後のデータ並べ替えを行った結果を示している。距離行列の対角線上の左上に、低分子化合物を含むクラスター1702が存在する。このクラスターの観察から、低分子化合物は、タンパク質の残基番号1、5、6のアミノ酸に近接していることがわかる。タンパク質-低分子化合物複合体のモデル1703に示すように、低分子化合物が残基番号の離れたタンパク質残基と隣接することは非常に多い。従来のDistance Matrix Plotでは、ポリペプチド鎖に沿ったクラスターの観察は容易だが、ポリペプチド鎖に沿わないが空間的に近いクラスターの同定は容易ではない。本実施例における方法では、上記で示したようにポリペプチド鎖に沿わないが空間的に近いクラスターの同定が、非常に容易になる。

さらに、蛋白質と低分子化合物の複合体における一部分を拡大表示したいときは、データ表示形式を変更して、各蛋白質残基の C_{α} 原子および低分子化合物の重心を原子間距離計算に使うかわりに、各蛋白質および低分子化合物を構成する全原子間の距離を用いることができる。もちろん、全原子間距離の計算から水素原子を省略してもよい。全原子表示においては、低分子化合物のどの原子と、蛋白質のどの残基内のどの原子が水素結合しているかを容易に見ることができる。

また、この方法を用いると、ある一つのタンパク質と一部異なる複数の低分子化合物との間のドッキング結果を表示するような場合、低分子化合物中の原子とタンパク質中の原子のどれが近接しているかを、複数のドッキング構造の間で比較することが、一つの行列の中で可能である。従来法の三次元的な構造図によって比較した場合は、慣れた研究者が時間をかけて図を観察する必要があるが、本実施例によれば、多くのドッキング構造間の比較を一瞥のもとに容易に、かつ定量的に行うことが可能になる。

産業上の利用可能性

二つの生体関連事象間の相関データを行列形式で表示する可視化方法に

において、本発明による可視化方法と、該可視化方法を実装したインターフェースを用いれば、相関データ規模の大小によって、相関データパターンの粗視化や、セル毎の情報の他ソースへのアクセス等の作業をマニュアルで実施することなく、相関データパターンとパターンを構成するセルに関する情報を、データ数の規模の変動に応じて自動的に選択された適切な表示形式と要約度で、同時に観察することが可能になる。これによって、表示すべきデータ数にかかわらず、個別セル内から得られる情報量を自動的に最大に保ちつつ、データの全体像の観察が可能になる。その結果、相関データの全体としての観察と少数データの詳細な観察を交互に繰り返す作業を、従来のマニュアルに比べ大幅に効率的に実施することが可能になり、大量の相関データからの有効な知識の発見を効率的に行うことが可能になる。

本発明を生体関連事象間の相互作用データ、例えばタンパク質－低分子化合物間相互作用データ、に適用したとき、利用者はこれらの相互作用の強度のすべてを一望に見ることができる。また、相互作用強度が類似したタンパク質や低分子化合物はデータの数が多いときには、データ量がコンパクトにまとめられた形で画面上に提示される。逆に、利用者が相互作用データのある一部に着目するときは、詳細な情報を閲覧しながら創薬研究における決定を下すことができる。タンパク質－タンパク質相互作用や、他の重要な相互作用データについても同様に本発明を用いて可視化しながら解析することで、創薬のプロセスにおけるデータ処理を加速し、ひいては創薬のスピードアップにつながる。

請 求 の 範 囲

1. 二つの生体関連事象間の相関データあるいは該相関データとそれぞれの事象の特徴データを行列形式で表示する可視化方法において、同一種類または異なる種類の生体関連事象間の相関データあるいは該相関データと各生体関連事象の特徴データを、所望する表示データのデータ数に応じて、予め用意された（a）複数のデータ表示形式から手動または自動的に選択された一つの形式と、（b）複数のデータ要約度から手動または自動的に選択された一つの要約度に基づいて画面表示することを特徴とする生体関連事象間相関データの可視化方法。

2. 前記（a）複数のデータ表示形式として、（A）一对の事象間の相関データを一つの表示データ単位とする表形式のデータ表示形式、（B）事象をクラスタリングした結果得られたクラスター間の相関データを一つの表示データ単位とする表形式のデータ表示形式、及び（C）相関データの集合を統計処理した結果を一つの表示データ単位とするデータの表示形式から選択される表示形式を用いることを特徴とする請求の範囲第1項に記載の可視化方法。

3. 前記（B）のクラスタリング方法として、二つの生体関連事象に関する属性情報、又は二つの生体関連事象間の相関情報に基づくクラスタリングを用いることを特徴とする請求の範囲第2項に記載の生体関連事象間相関データの可視化方法。

4. 前記（B）のクラスター間の相関データを一つの表示データ単位とする表形式のデータ表示形式において、相関強度が強いクラスターから順番に表の左上から順番に対角線上に結果を並べ替える機能を持つことを特徴とする請求の範囲第2項に記載の可視化方法。

5. 前記（b）複数のデータ要約度として、データフィールドの表示または非表示、文字型のデータフィールド中のデータの短縮、及び数値型データフィールド中のデータの短縮から選択される要約方法を用いることを特徴とする請求の範囲第1項に記載の可視化方法。

6. 前記文字型のデータフィールド中のデータの短縮が、階層構造を有する文字情報から該階層の一部分を抽出する操作、文字データ中からあらかじめ登録されているキーワードを抽出する操作、及び文字データを一つの記号や文字または色彩で対応させる操作からなることを特徴とする請求の範囲第3項に記載の可視化方法。

7. 前記数値型のデータフィールド中のデータの短縮が、数値を任意の有効数字で丸める操作、数値の指数部分のみを取り出す操作、及び一定範囲の数値を色彩で対応させる操作からなることを特徴とする請求の範囲第3項に記載の可視化方法。

8. 画面表示形式とデータの要約度の自動的な選択方法として、画面表示すべき相関データのエントリー数及びあらかじめ指定された情報表示領域と情報表示単位のサイズに応じて、最大の情報量を与えるデータ表示形式とデータ要約度の組を選択することを特徴とする請求の範囲第1項に記載の可視化方法。

9. 前記生体関連事象間の複数種類の相関データを、行列のセル中に同時に識別表示することを特徴とする請求の範囲第1項に記載の可視化方法。

10. 前記生体関連事象間の相関データが、低分子化合物とタンパク質の相互作用であることを特徴とする請求の範囲第1項に記載の可視化方法。

11. 前記生体関連事象として、一つ以上の分子の複合体において、各分子毎に分子内原子、または分子内原子の集合から構造単位を定義し、該構造単位を構成する原子の座標から該構造単位の代表位置を定義し、該構造単位の各々を行及び列の要素として持ち、該構造単位の代表位置間の距離情報を行列のセル中表示することを特徴とする請求の範囲第1項に記載の生体関連事象間相関データの可視化方法。

12. 請求の範囲第2項におけるクラスター内の、メンバー間に共通の生体関連事象の特徴量を抽出することを特徴とする二つの生体関連事象間相関情報の解析方法。

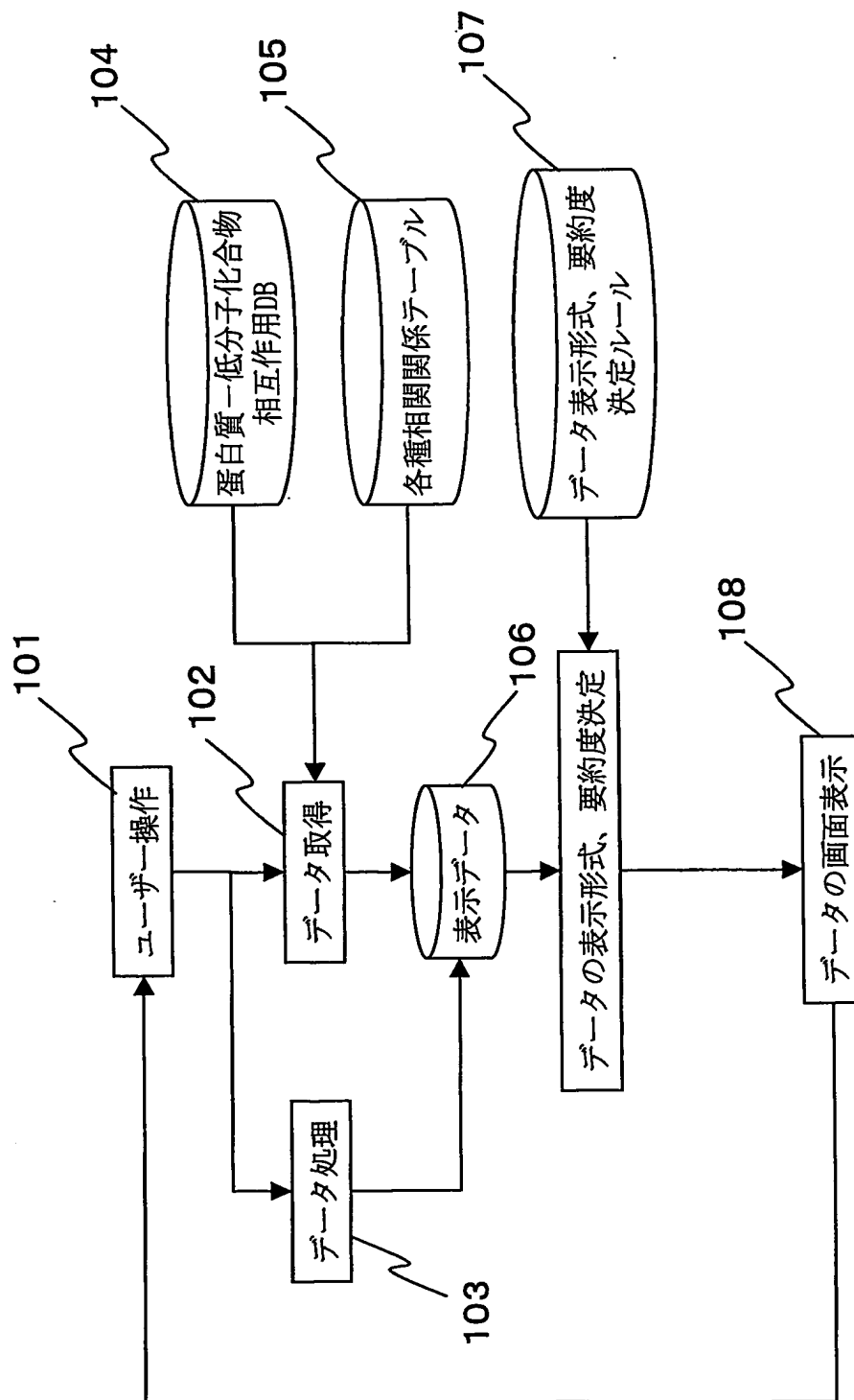
13. 前記生体関連事象の特徴量が、数値又はテキストから成る一つの要素、又は複数の要素、あるいは分子の三次元構造を表す特徴量で表現さ

れることを特徴とする請求の範囲第 1 2 項に記載の二つの生体関連事象間
相関情報の解析方法。

1 4. 請求の範囲第 1 2 項又は第 1 3 項に記載の解析方法によって得ら
れたクラスターの解析結果を、非特許文献や特許文献から抽出された関連
既知情報と共に格納し、クラスター解析結果からの既知関連情報の検索や、
既知情報からのクラスター解析結果の検索機能を有したデータベース。

1 5. 請求の範囲第 1 項～第 1 4 項に記載の可視化方法、解析方法、デ
ータベースをコンピューターに実行させるためのプログラムを記録したコ
ンピューター読み取り可能な記録媒体。

第1図



第2図

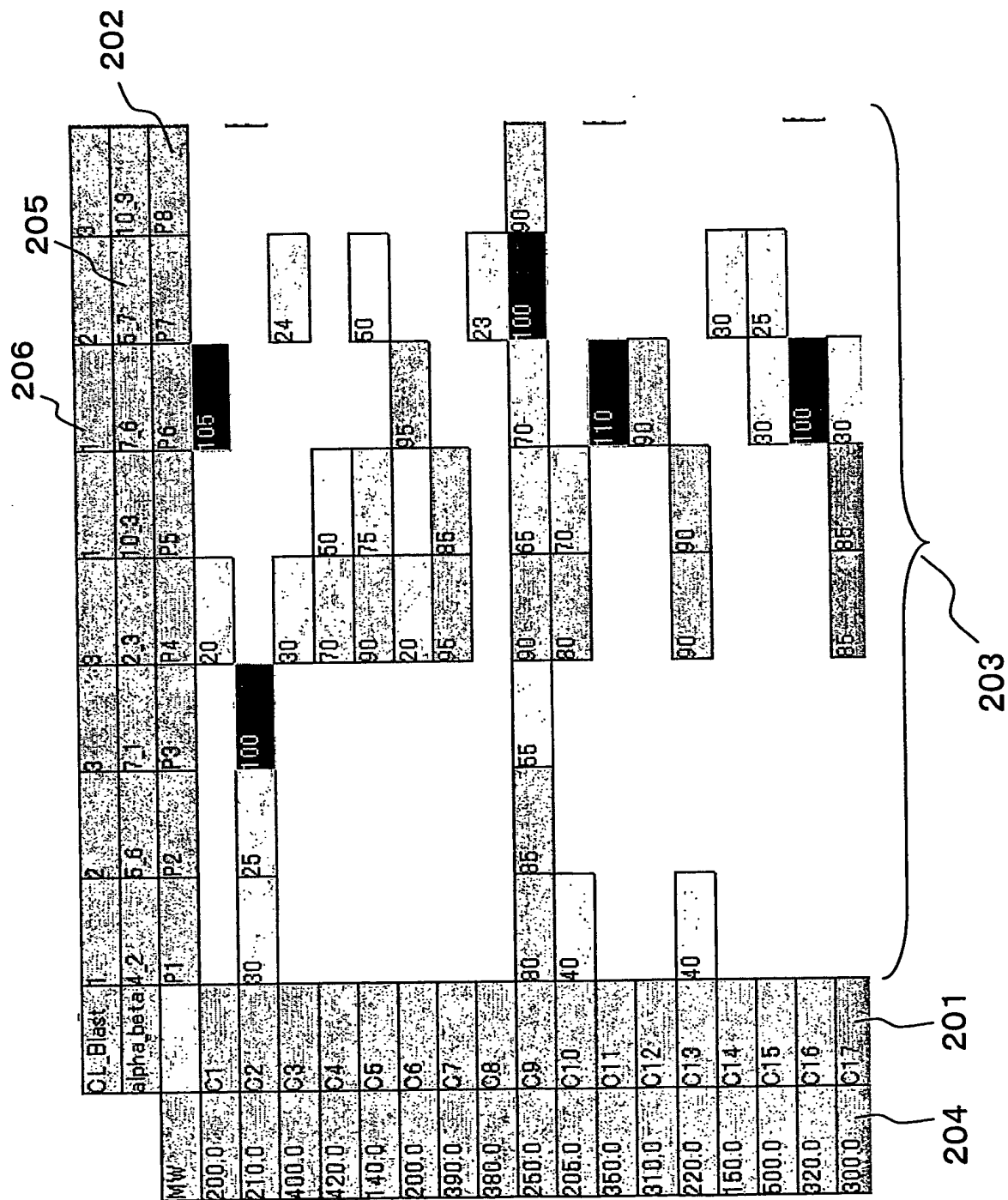
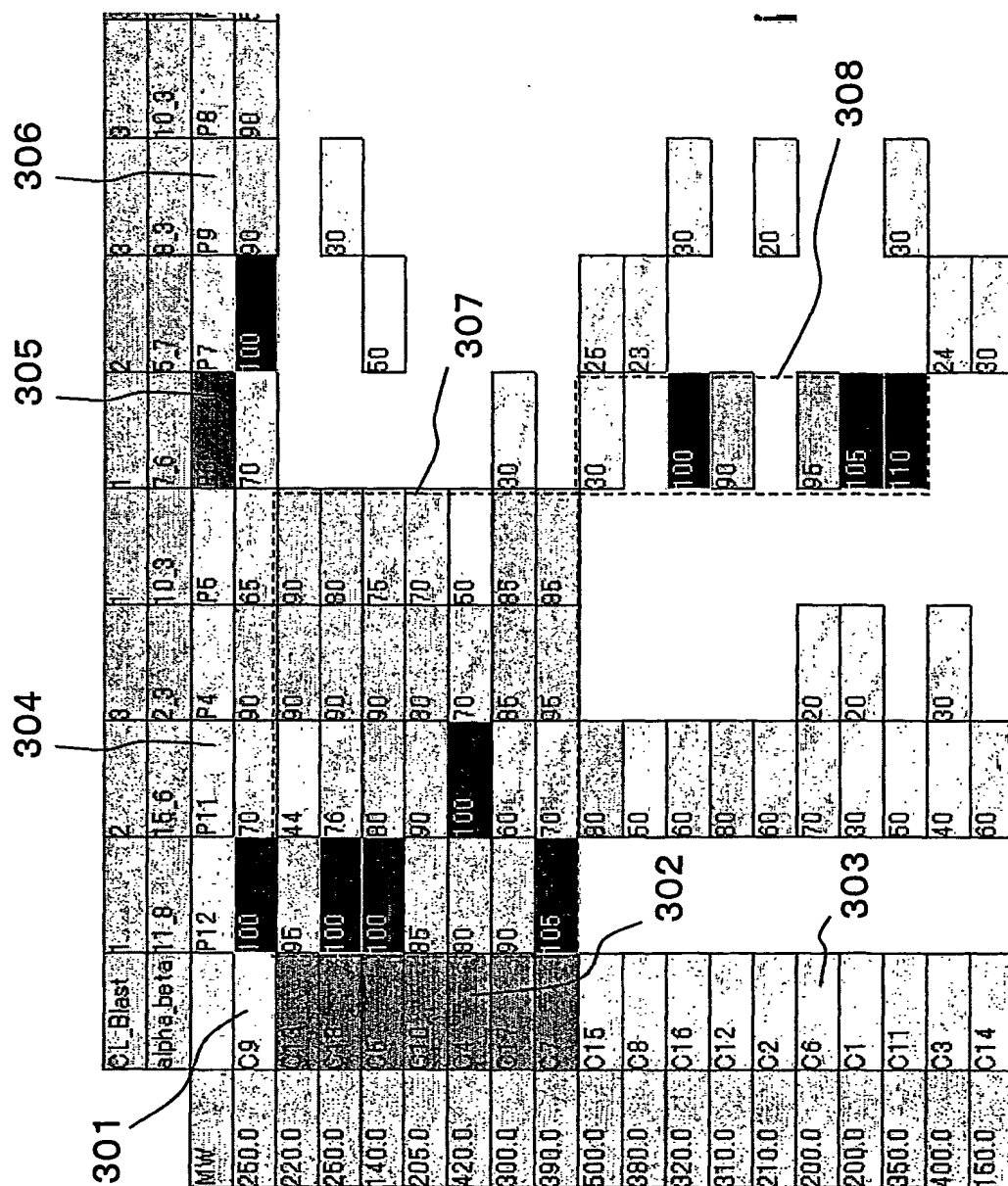
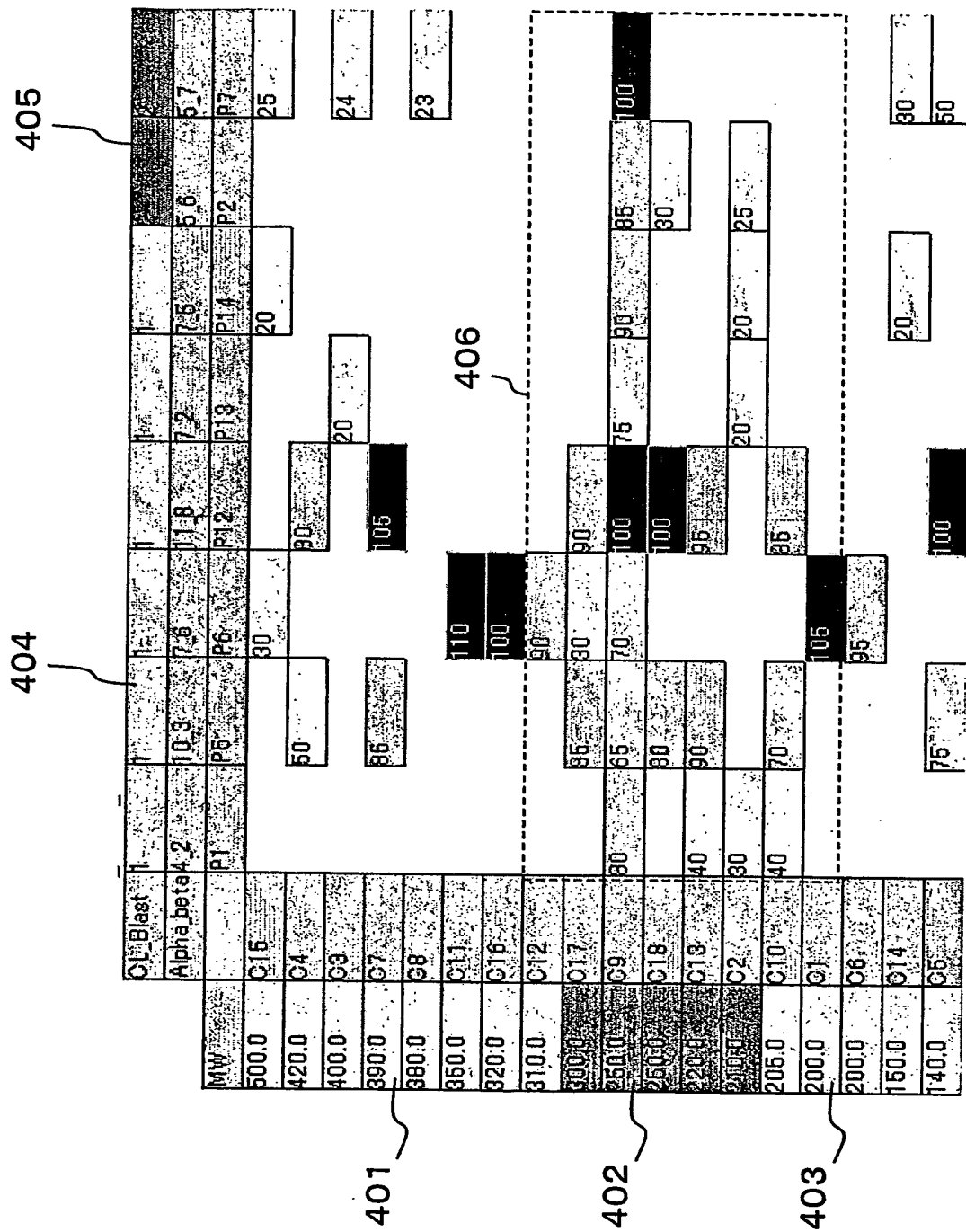


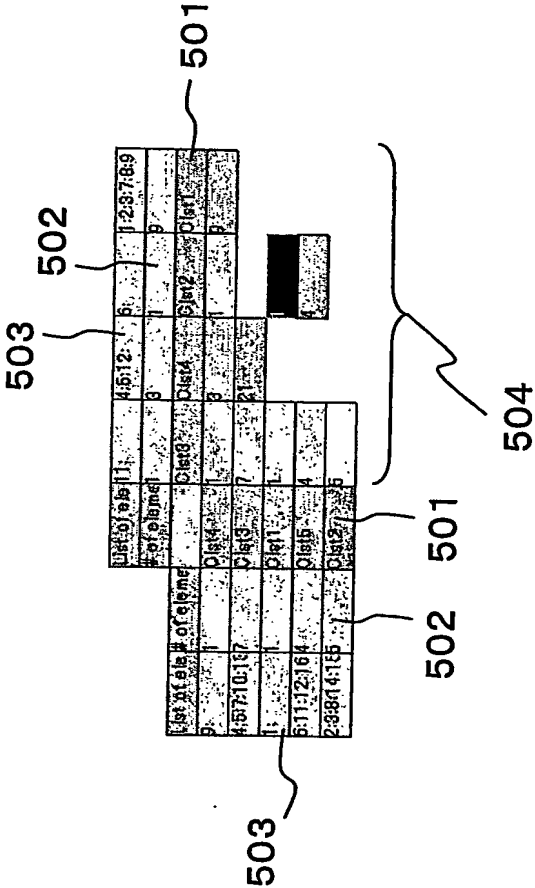
図 3
鋸



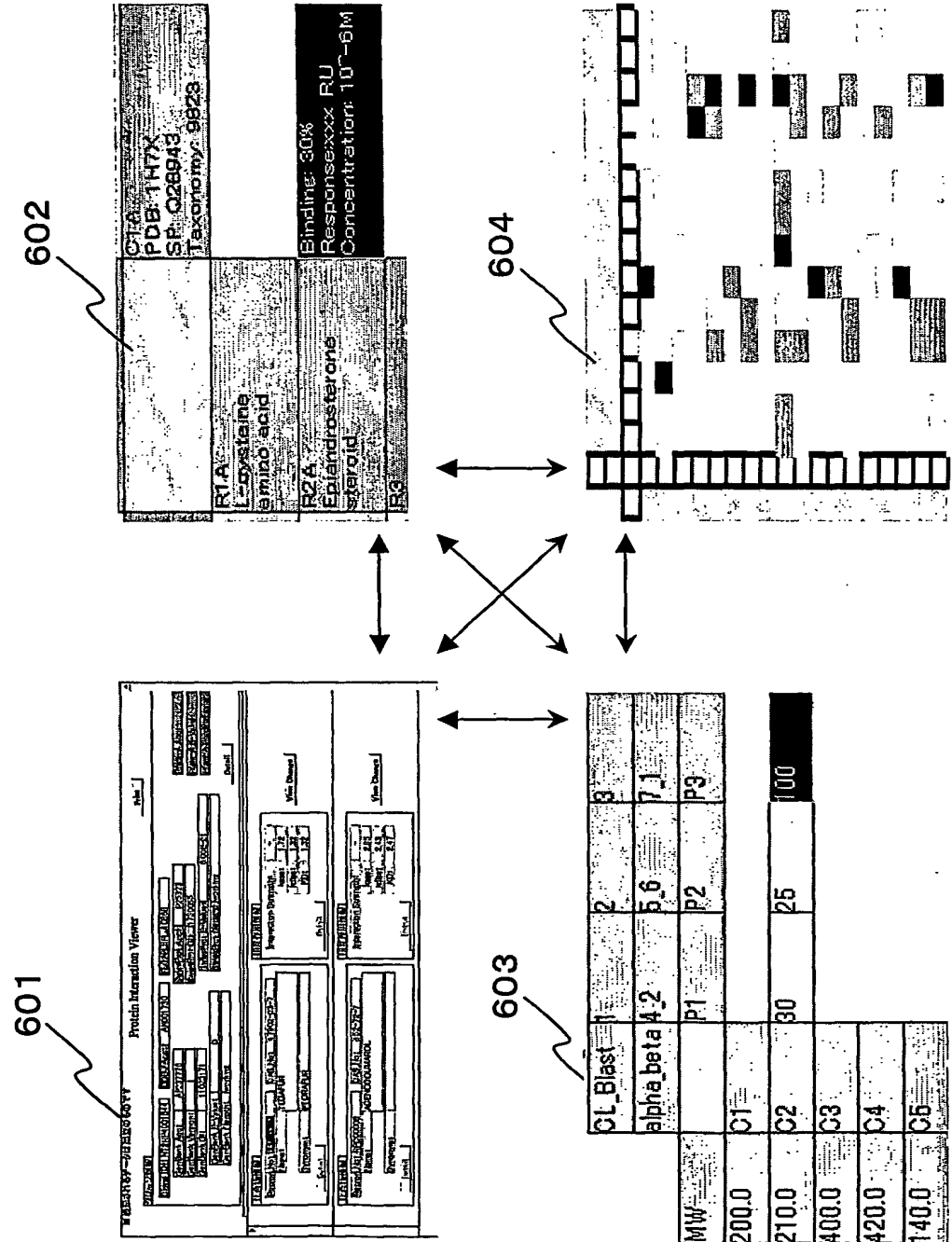
第4図



第5図



第6図



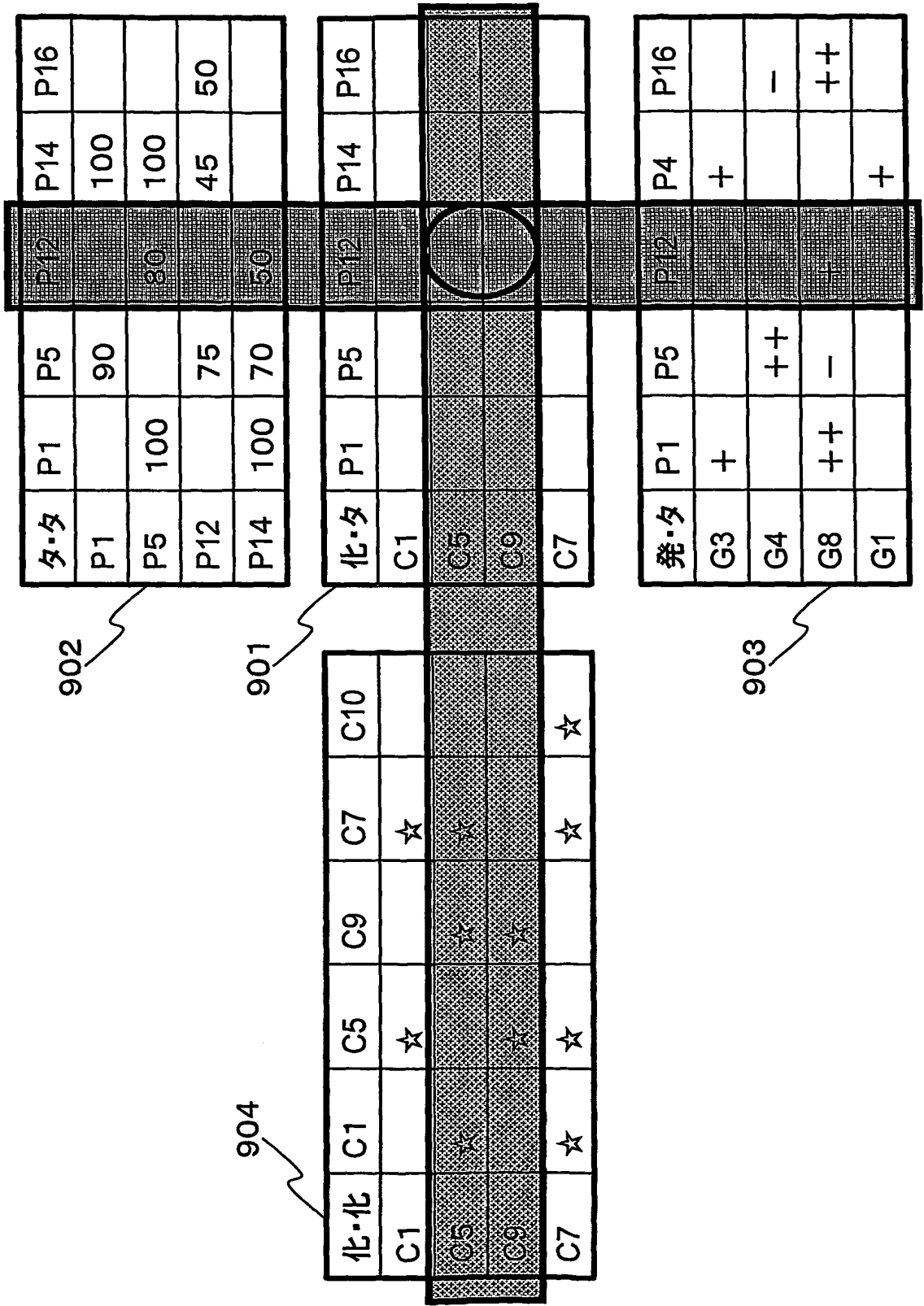
第7図

要約度	データ項目	場所	要約ルール
1	化合物名	ラベル	そのまま 705
1	分子量	特徴量記載セル	そのまま
1	物性クラスター番号	特徴量記載セル	そのまま
1	薬効クラスター名	特徴量記載セル	そのまま
2	化合物ID	ラベル	下5桁
2	分子量	特徴量記載セル	整数化
2	物性クラスター番号	特徴量記載セル	下5桁
2	薬効クラスター番号	特徴量記載セル	下5桁
2	薬効クラスター名	情報表示別画面	薬効クラスター番号からのリンク。そのまま
3	化合物ID	情報表示別画面	ラベルからのリンク。そのまま 706
3	分子量	特徴量記載セル	色彩(200、300、400、500)
3	物性クラスター番号	特徴量記載セル	色彩(番号ごとに違う色)
3	薬効クラスター番号	特徴量記載セル	色彩(番号ごとに違う色)

第8図

条件	801	表示形式	802	要約度	803
$P \times C \leq 3$		個々データ表示		0	
$G \leq 11$ & $R \leq 11$		個々データ表示		1	
$G \leq 34$ & $R \leq 22$		個々データ表示		2	
$G \leq 135$ & $R \leq 270$		個々データ表示		3	
$G_c \leq 11$ & $R_c \leq 11$		クラスター表示		1	
$G_c \leq 34$ & $R_c \leq 22$		クラスター表示		2	
$G_c \leq 135$ & $R_c \leq 270$		クラスター表示		3	
上記以外		統計量表示		0	
定 義	G: P+列方向物性表示数+1 R: C+行方向物性表示数+1 Gc: Pc+列方向物性表示数+1 Rc: Cc+行方向物性表示数+1				

第9図

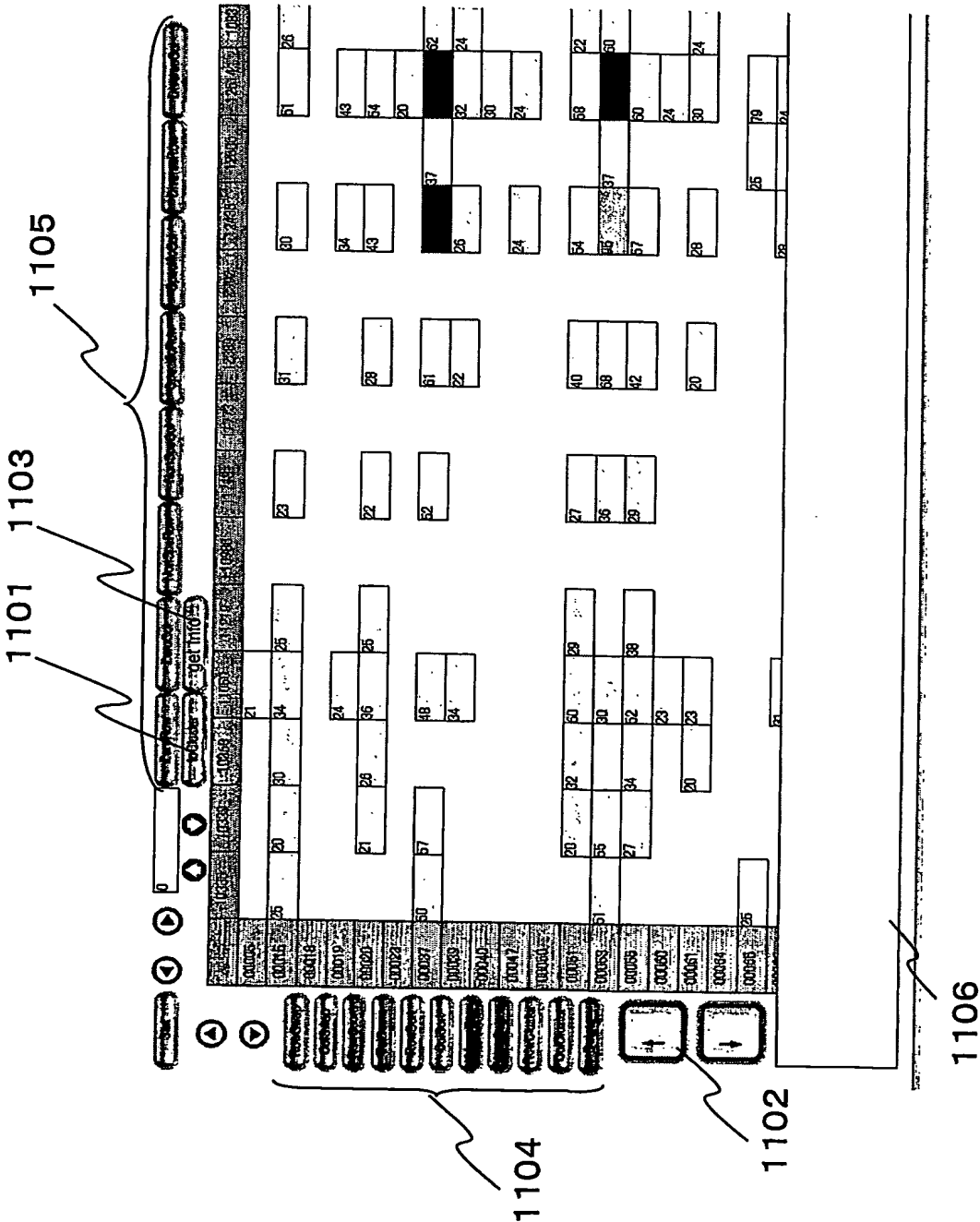


第10図

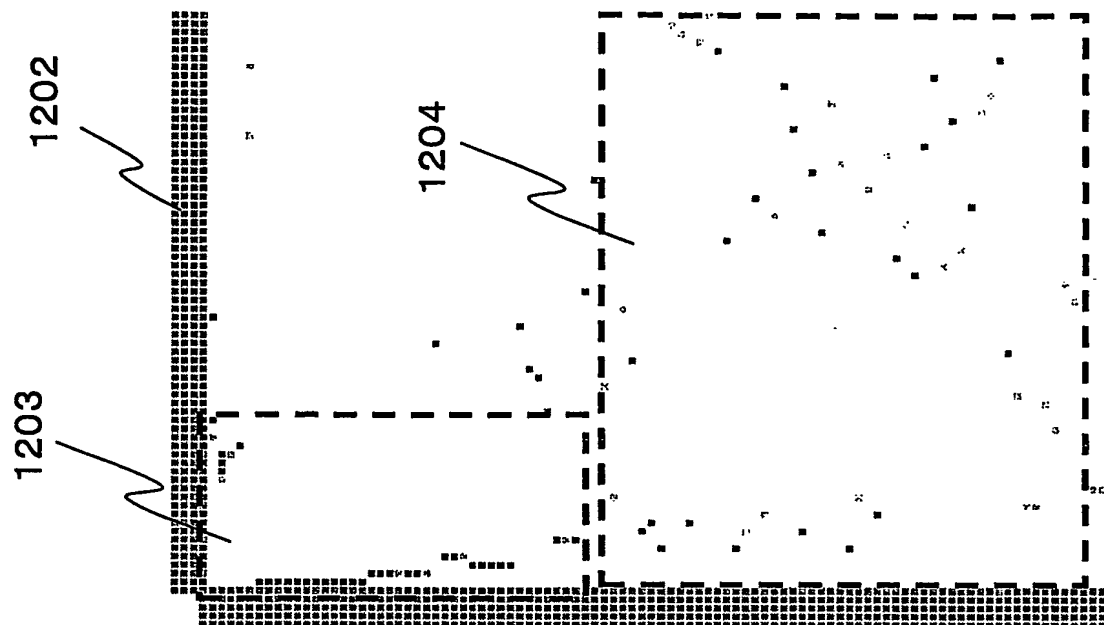
(R5, C12), (R9, C12)の関連情報
 タンパク質—発現テ—ブルより:
 タンパク質—タンパク質相互作用テ—ブルより:
 低分子化合物—低分子化合物相互作用テ—ブルより:

xx件
yy件
zz件

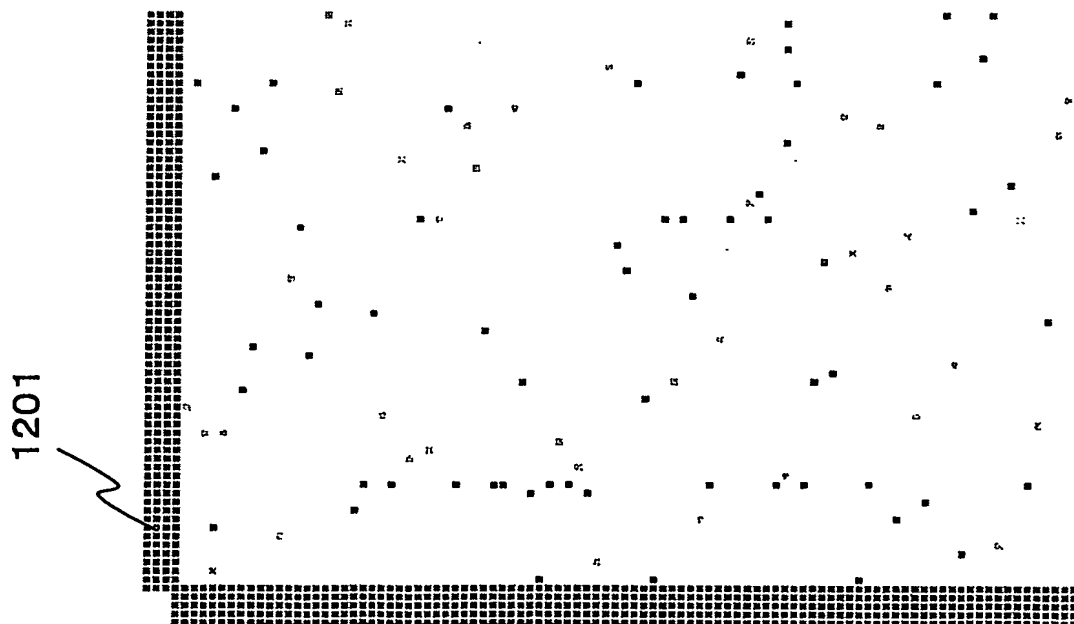
第11図



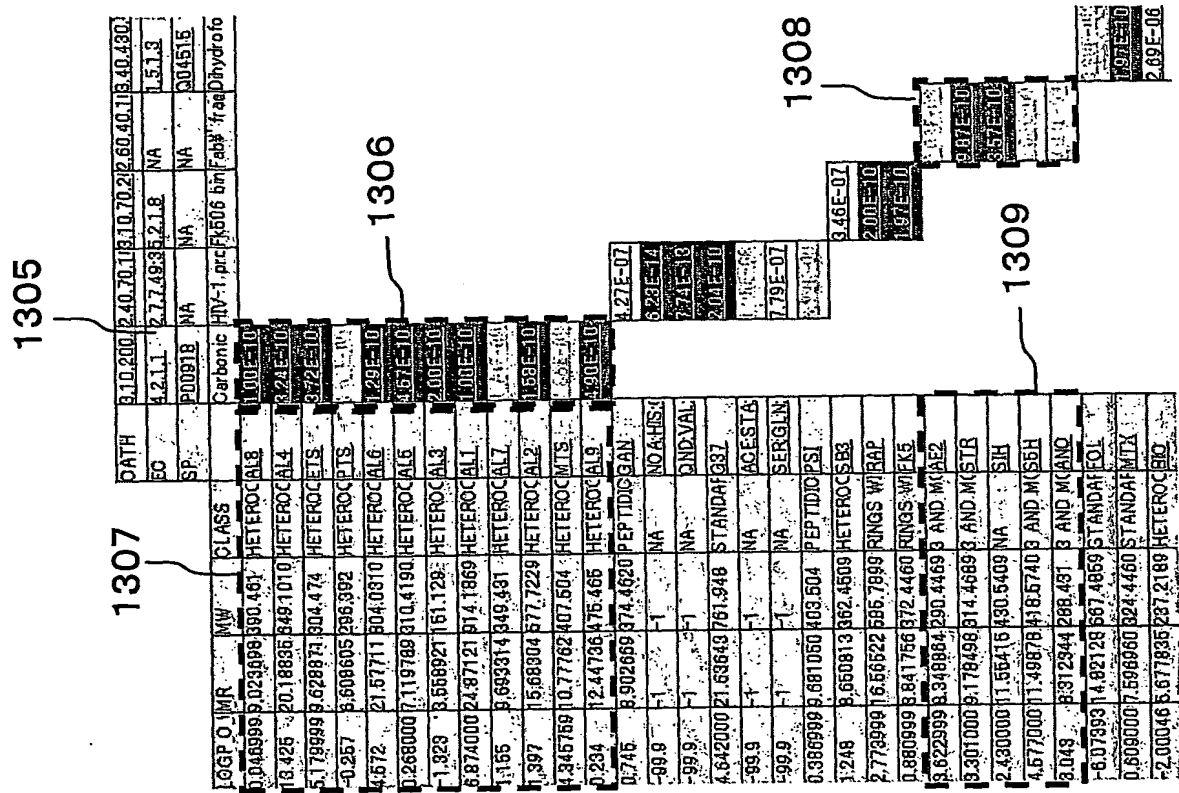
第12B図



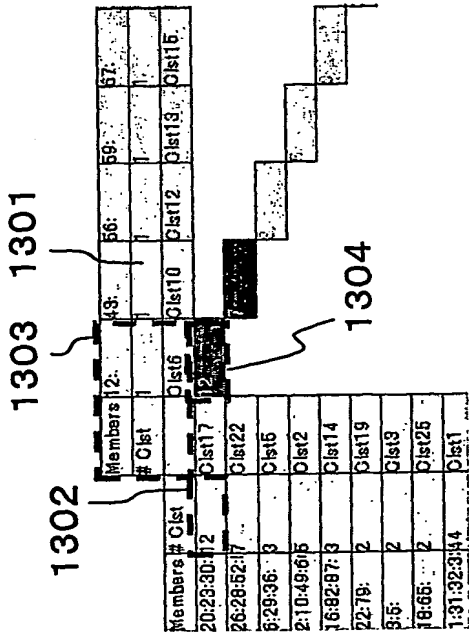
第12A図



第13B図



第13A図



第13C図

1310

LogP	MR	化合物名	結合
中	小	AE2	弱
小	中	STR	強
小	大	SIH	強
大	大	S5H	中
小	小	ANO	中

第14A図

第14B図

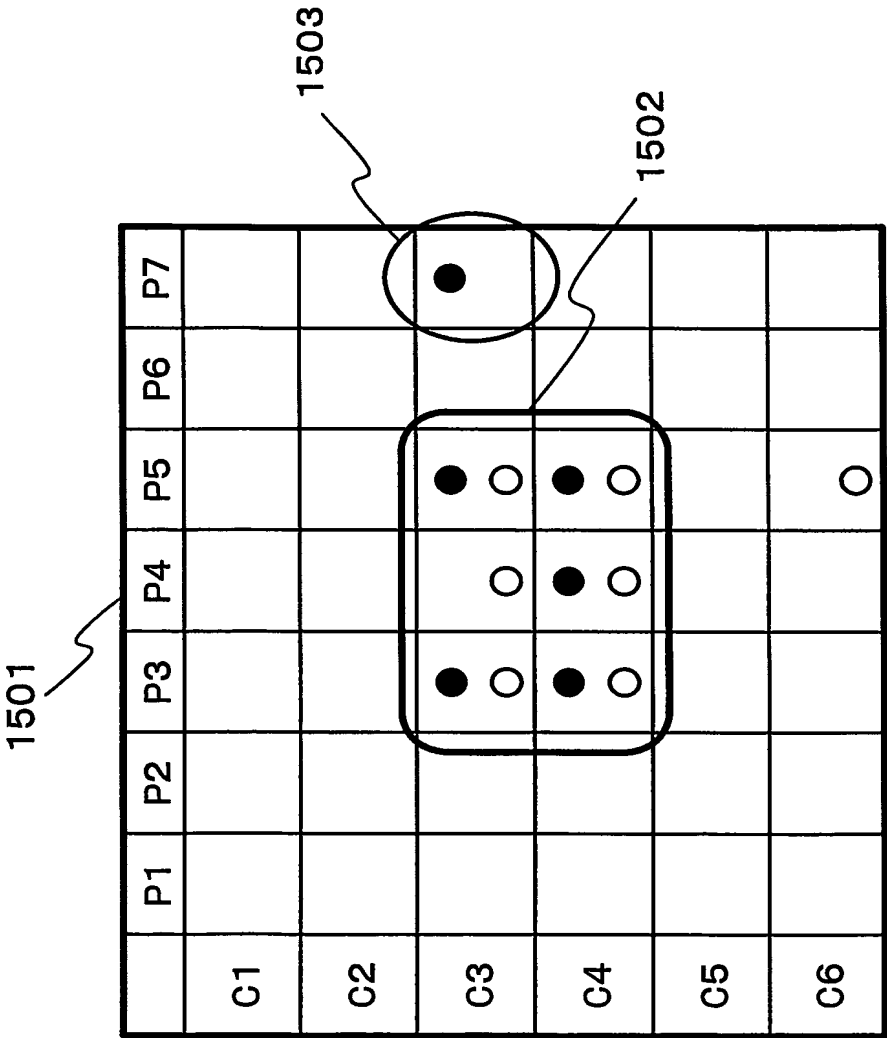
	P1	P2	P3	P4	P5	P6	P7
T1		●				●	●
T2				●	●		
T3			●	●	●		
T4				●	●	●	●
T5						●	●
T6	●			●	●		
T7			●				

	S1	S2	S3	S4	S5
C1	●				
C2		●	●		
C3		●	●		
C4		●	●	●	●
C5		●	●	●	●
C6			●		

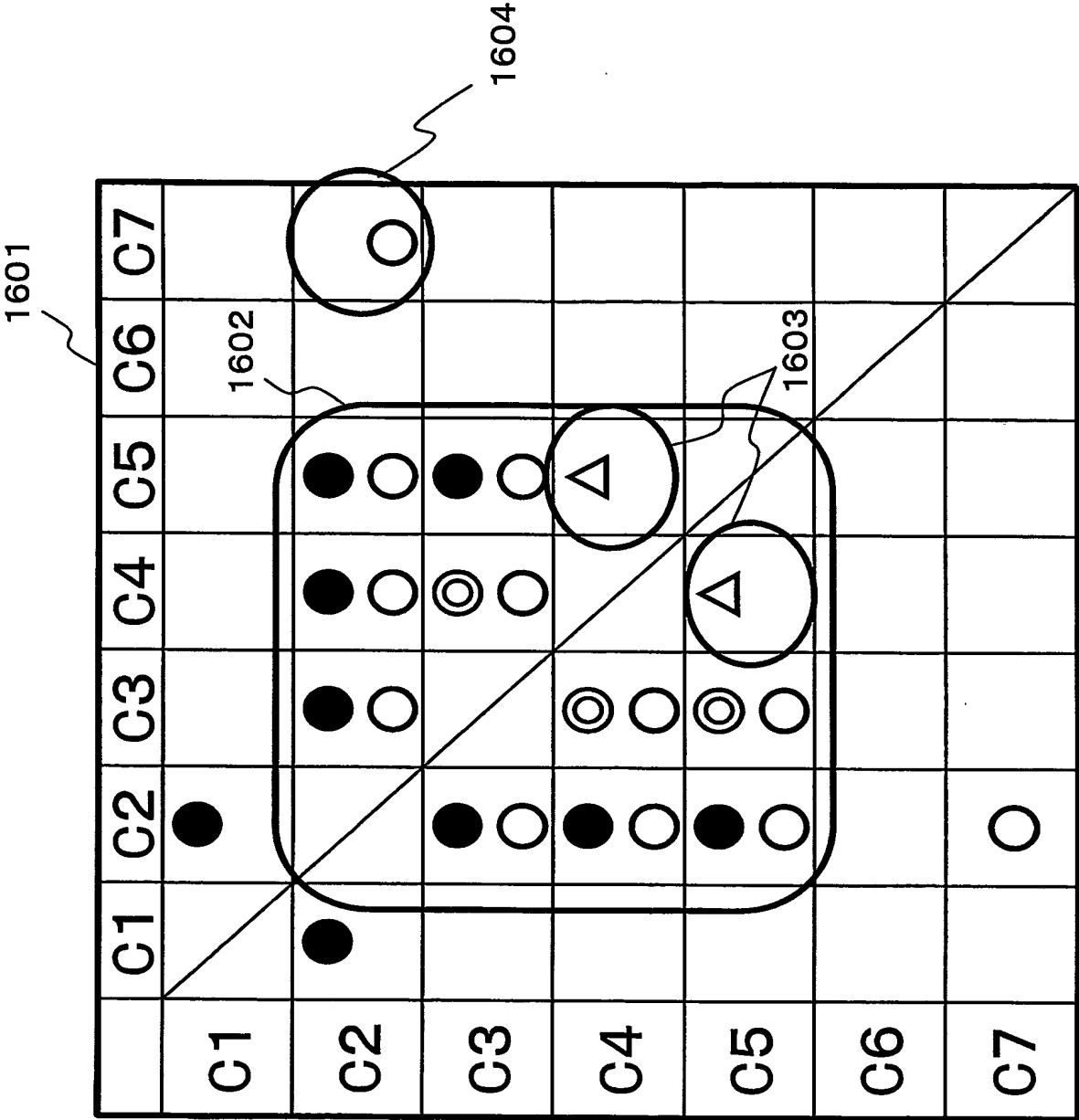
第14C図

	P1	P2	P3	P4	P5	P6	P7
C1							
C2	●	●	●	●	●	●	●
C3	●	●	●	●	●	●	●
C4	●	●	●	●	●	●	●
C5	●	●	●	●	●	●	●
C6							

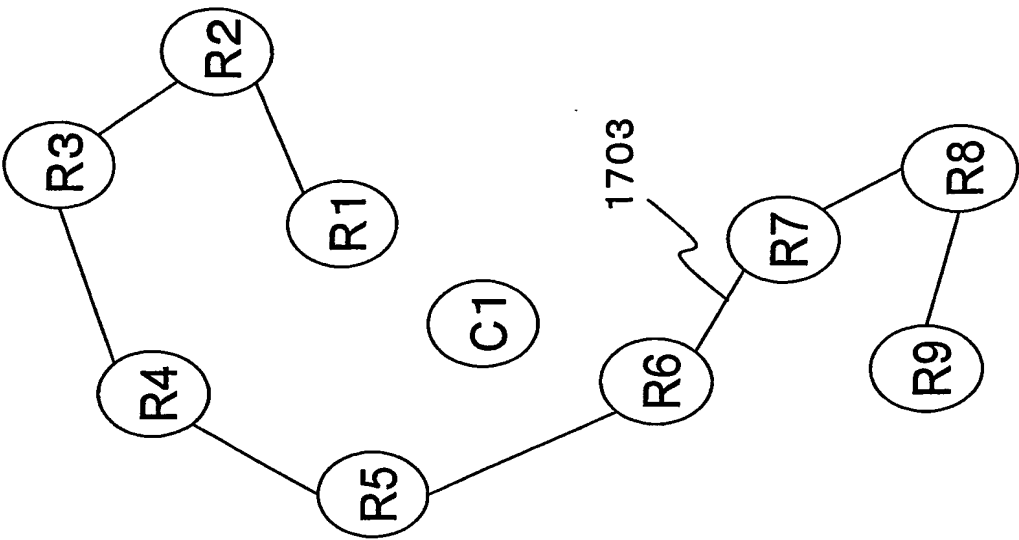
第15図



第16図



第17B図



第17A図

1701

	R3	R2	R4	R1	R5	C1	R6	R7	R9	R8
R3		●								
R2	●			●						
R4		●		●	●					
R1			●	●	●	●	●	●	●	●
R5			●	●	●	●	●	●	●	●
C1				●	●	●	●	●	●	●
R6					●	●	●	●	●	●
R7							●	●	●	●
R9								●	●	●
R8										●

1702

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/010250

A. CLASSIFICATION OF SUBJECT MATTER

Int.Cl⁷ G06F19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Int.Cl⁷ G06F19/00, G06F3/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho 1922-1996 Toroku Jitsuyo Shinan Koho 1994-2004

Kokai Jitsuyo Shinan Koho 1971-2004 Jitsuyo Shinan Toroku Koho 1996-2004

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JICST FILE (JOIS)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Hiroaki KITANO, "System Biology Seimei o System toshite Rikai suru", Shujunsha Co., Ltd., 01 July, 2001 (01.07.01), pages 72 to 90	1-15
A	JP 10-269050 A (Canon Inc.), 09 October, 1998 (09.10.98), Column 32, line 23 to column 33, line 5; Fig. 16 & EP 0867824 A2 & US 2003/0030663 A1	1-15
A	Hidekazu KAWASHIMA, "Shinseiki Iryo o Mezashite -SNP to DNA Chip DNA Chip to Bioinformatics", Gene & medicine, Kabushiki Kaisha Medical Do, 10 February, 2000 (10.02.00), Vol.4, No.1, pages 129 to 133	1-15

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
08 October, 2004 (08.10.04)

Date of mailing of the international search report
02 November, 2004 (02.11.04)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2004/010250

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2003-505749 A (Warner-Lambert Co.), 12 February, 2003 (12.02.03), Full text; Figs. 1 to 10 & WO 2000/004338 A1	1-15
A	JP 2003-242154 A (Celestar Lexico-Sciences, Inc.), 29 August, 2003 (29.08.03), Full text; Figs. 1 to 19 & WO 2003/081471 A1	1-15
A	JP 2002-149300 A (Takumi HIGASHIBARA), 24 May, 2002 (24.05.02), Full text; Figs. 1 to 9 & WO 2000/004338 A1	1-15
A	JP 11-085448 A (Matsushita Electric Industrial Co., Ltd.), 30 March, 1999 (30.03.99), Full text; Figs. 1 to 55 & EP 0901084 A2 & US 6246442 B1	1-15

A. 発明の属する分野の分類 (国際特許分類 (IPC)) <p style="text-align: center;">Int. Cl⁷ G06F19/00</p>		
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) <p style="text-align: center;">Int. Cl⁷ G06F19/00 G06F3/00</p>		
最小限資料以外の資料で調査を行った分野に含まれるもの <div style="display: flex; justify-content: space-between;"> <div>日本国実用新案公報</div> <div>1922-1996年</div> </div> <div style="display: flex; justify-content: space-between;"> <div>日本国公開実用新案公報</div> <div>1971-2004年</div> </div> <div style="display: flex; justify-content: space-between;"> <div>日本国登録実用新案公報</div> <div>1994-2004年</div> </div> <div style="display: flex; justify-content: space-between;"> <div>日本国実用新案登録公報</div> <div>1996-2004年</div> </div>		
国際調査で使用した電子データベース (データベースの名称、調査に使用した用語) <p style="text-align: center;">JICSTファイル (JOIS)</p>		
C. 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	北野宏明, システムバイオロジー 生命をシステムとして理解する, 株式会社秀潤社, 2001. 07. 01, p. 72-90	1-15
A	JP 10-269050 A (キヤノン株式会社) 1998. 10. 09, 第32欄第23行-第33欄第5行, 第16図 & EP0867824 A2 & US 2003/0030663. A1	1-15
<input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。		
* 引用文献のカテゴリー <div style="display: flex;"> <div style="flex: 1;"> <p>「A」 特に関連のある文献ではなく、一般的技術水準を示すもの</p> <p>「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの</p> <p>「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)</p> <p>「O」 口頭による開示、使用、展示等に言及する文献</p> <p>「P」 国際出願日前で、かつ優先権の主張の基礎となる出願</p> </div> <div style="flex: 1;"> <p>の日の後に公表された文献</p> <p>「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの</p> <p>「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの</p> <p>「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの</p> <p>「&」 同一パテントファミリー文献</p> </div> </div>		
国際調査を完了した日 <p style="text-align: center;">08. 10. 2004</p>		国際調査報告の発送日 <p style="text-align: center; font-size: 1.2em;">02.11.2004</p>
国際調査機関の名称及びあて先 日本国特許庁 (ISA/J P) 郵便番号 100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 石川 正二 <div style="display: flex; justify-content: space-between; margin-top: 5px;"> <div>電話番号 03-3581-1101</div> <div>内線 3560</div> </div>

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	川島秀一, “新世紀医療をめざして—SNPとDNAチップ DN Aチップとバイオインフォマティクス”, 遺伝子医学, 株式会社メ ディカルドゥ, 2000. 02. 10, 第4巻, 第1号, p. 129-133	1-15
A	JP 2003-505749 A (ワーナー・ランバート カン パニー) 2003. 02. 12, 全文, 第1-10図 & WO 2000/004338 A1	1-15
A	JP 2003-242154 A (セレスター・レキシコ・サイ エンシズ株式会社) 2003. 08. 29, 全文, 第1-19図 & WO 2003/081471 A1	1-15
A	JP 2002-149300 A (東原功) 2002. 05. 2 4, 全文, 第1-9図 & WO 2000/004338 A1	1-15
A	JP 11-085448 A (松下電器産業株式会社) 199 9. 03. 30, 全文, 第1-55図 & EP 0901084 A2 & US 6246442 B1	1-15

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.